V. G. Gorbatskii

EXPLODING STARS AND GALAXIES

Translated from Russian

Published for the National Aeronautics and Space Administration
and the National Science Foundation, Washington, D.C.
by the Israel Program for Scientific Translations

V.G. Gorbatskii

EXPLODING STARS AND GALAXIES

(Kosmicheskie vzryvy)

Izdatel'stvo "Nauka"
Moskva 1967

Translated from Russian

Israel Program for Scientific Translations
Jerusalem 1970

NASA TT F-559
TT 69-55022

Published Pursuant to an Agreement with
THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
and
THE NATIONAL SCIENCE FOUNDATION, WASHINGTON, D. C.

IPST Cat. No. 5440
UDC 523

Translated by Z. Lerman

Printed in Jerusalem by Keter Press
Binding: Wiener Bindery Ltd., Jerusalem

Available from the
U. S. DEPARTMENT OF COMMERCE
Clearinghouse for Federal Scientific and Technical Information
Springfield, Va. 22151

XII/16/3

Table of Contents

FOREWORD

Our understanding of the Universe has greatly expanded in recent years. Among the most significant topics tackled by the modern tools of astronomy and astrophysics were the explosions in galactic nuclei — cosmic catastrophes of extraordinary power — and quasars — extremely distant objects of exceptionally high luminosity. These phenomena attracted the attention of wide circles of scientists, and the repercussions of these discoveries were not limited to physics and astrophysics.

Explosive processes are a common phenomenon in the stellar universe. Recent discoveries only confirmed the important role of explosions in the evolution of stars and giant star systems. So far, however, no account of the whole range of cosmic explosions — from solar flares to exploding galactic nuclei — has been published for the layman. The aim of our book is to fill this gap to a certain extent.

The physics of cosmic explosions is treated in considerable detail. Various numerical estimates are made with the aim of ensuring a more complete understanding of the effects involved. The simple formulas used for these estimates are mostly taught as part of the elementary physics course in high school, and whenever more sophisticated formulas are required, their derivation is based on simple reasoning. Although often the argument is not fully rigorous, it always reflects the physics of the problem.

The explanation of the various effects considered in this book is one of the most difficult problems of modern physics. The reader therefore cannot expect to read through even this superficial account without minimum effort. It is the author's hope, however, that the reader will be rewarded for his trouble with a more complete and thorough understanding of the world we live in.

The author's intention was to make the book accessible to any layman with high-school education. Clearly, a reader familiar with the various techniques used in the derivations will be able to assimilate the results more readily. We have therefore devoted fairly much attention and effort to explaining the experimental and theoretical methods used, although the limited scope of the book imposed an obvious restriction on the amount of detail that could be put in. Fortunately, the reader can continue the study of the subject with the aid of any of the excellent books listed at the end, following the Conclusion.

Some of topics considered in the book have been so far treated in scientific papers only. This includes, for example, the role of explosions in the evolution of celestial bodies. For this reason the book may prove of some interest for the more qualified readers too.

The author would like to acknowledge the assistance of S. A. Kaplan and V. V. Ivanov, who read the manuscript and made numerous valuable comments. G. B. Gel'freikh helped with kind advice in connection with §5 of the book, and E. A. Dibai discussed §8 and §9.

§ 1. WHAT IS AN EXPLOSION?

Explosive processes are a very common occurrence in nature observed in a variety of cases, so there can hardly be anybody not familiar with the basic features of explosions. Superficial acquaintance with the phenomenon, however, is not sufficient for understanding such complex effects as explosions of stars. First, what we call an explosion can be caused by a wide range of factors. Thus a steam boiler explodes when the superheating exceeds a certain permissible critical value, and an artillery shell explodes following the combustion of the explosive in some chemical reaction. Second, explosions show a great variety of external characteristics. The electric discharge, for instance, is superficially dissimilar to an earthquake. Therefore, before proceeding with a discussion of cosmic explosions, we should elucidate the essential features of the phenomenon and establish exactly what processes are to be classified as explosions.

An explosion is generally regarded as a certain physical state of a body or a system which involves a very rapid, virtually catastrophic conversion of energy from one form to another and its release into the ambient medium. If we use this definition, explosions are seen to include such phenomena as a nuclear explosion and a lightning discharge, the impacting of a meteorite on the Earth and the bursting of a gas filled balloon, and so on.

A necessary condition of an explosion is the fast conversion of energy from one form to another. If this conversion is slow and protracted, the characteristic features of an explosion are not observed. For example, nuclear energy is continuously released in a reactor by various slow nuclear reactions, but this is not an explosion. If, however, the reaction is greatly accelerated and the energy normally released over a span of hours is released within a few seconds, we have an obvious explosion.

In this section we will study the general properties of explosions occurring under terrestrial conditions. These explosions entail the conversion of chemical, nuclear, electric, or mechanical energy, as the case may be, into heat and radiation. In some cases a substantial part of the energy released in the explosion is carried off by the fast particles produced by the explosion, protons electrons, and others. Fragments of the exploding object may also acquire a substantial kinetic energy as a result of the explosion.

A considerable proportion of the energy released by the explosion is converted into the kinetic energy of the surrounding medium; this is generally regarded as the most typical feature of an explosive process. When a bomb or a shell explode, the air is drastically expelled from the explosion focus. The medium acquires its kinetic energy mainly through heat, either released directly by the explosion or produced from other forms of free energy. To fully grasp the nature of explosive phenomena, we should therefore start with a discussion of the basic thermodynamic concepts.

What do we exactly mean when we say that heat is released in the process? The release of heat is evident from the increase in the temperature of the physical objects, whether solid, liquid, or gaseous, near the explosion focus. Any physical object is made up of individual particles — atoms or molecules. The temperature of the object provides a measure of the mean kinetic energy of the constituent particles. The higher the temperature, the faster the particles move. An explosion thus increases the kinetic energy of particles near the explosion focus. Thus, when powder explodes in a cartridge, the very fast chemical combustion reaction produces a hot gas. The chemical energy stored in the powder is converted into kinetic energy of gas particles.

Hot gas particles exchange energy when they collide with one another, but the total energy of all the gas particles in a certain enclosed volume does not change. The kinetic energy of each particle is equal to half the product of the particle mass by the square of its velocity. If we add up the kinetic energies of all the constituent particles and divide the sum by the number of particles, we obtain the mean particle energy $m\overline{v^2}/2$. Here $\overline{v^2}$ is the mean square velocity, and $\sqrt{\overline{v^2}}$ is called the root mean velocity.* Any gas always contains a certain proportion of particles with velocities much greater than the root mean square velocity, and also particles with velocities less than $\sqrt{\overline{v^2}}$.

The temperatures of objects are generally expressed in the so-called absolute scale (in degrees Kelvin, °K). The zero point on this scale corresponds to zero kinetic energy of particles. Experiments show that the velocity of gas molecules should decrease to zero as the temperature of the gas approaches $-273.16°$ on the centigrade scale. The zero point of the absolute Kelvin scale is therefore taken to coincide with the above temperature on the centigrade scale.

The absolute temperature T and the mean kinetic energy of particles are related by the equation

$$\frac{m\overline{v^2}}{2} = \frac{3}{2}kT, \tag{1}$$

where $3k/2$ is a proportionality coefficient. It is different from unity, since the degree divisions of the temperature scale are chosen from practical considerations. The coefficient k, known as the Boltzmann constant, is numerically equal to $1.38 \cdot 10^{-16}$ erg/deg in the CGSE system of units.

Equation (1) can be used to calculate the mean velocity of gas particles at different temperatures. Thus, at 15°C, which corresponds to 288°K, the mean square velocity of a nitrogen molecule (its mass is $46.5 \cdot 10^{-24}$ g**) is 467 m/sec.

* The mean square velocity in general is not equal to the mean (or average) velocity. Consider a system of two particles with velocities v_1 and v_2. Here the mean square velocity is $\overline{v^2} = \frac{v_1^2 + v_2^2}{2}$, whereas the square of the mean velocity is $(\overline{v})^2 = \left(\frac{v_1 + v_2}{2}\right)^2 = \frac{v_1^2 + v_2^2 + 2v_1v_2}{4}$. Generally the difference between the $\overline{v^2}$ and $\overline{v}^2$ of a gas is not very large.

** The mass of a molecule can be found by dividing the mass of one gram-molecule through the number of molecules (i.e., Avogadro's number, $6.0 \cdot 10^{23}$).

The molecules in a hot gas move in all directions; they strive to occupy as large a volume as possible, i.e., the gas expands. If the particles encounter an obstacle in a certain direction (for simplicity, consider a rigid wall as such an obstacle), they exert a definite pressure on that obstacle and try to move it. To keep the obstacle in a fixed position, a certain counterforce is required. The magnitude of this counterforce per unit wall area is defined as the pressure of the gas.

The most obvious outcomes of explosions, and in particular their destructive power, are associated with the exceedingly high pressures of the hot gases. We should therefore consider in some detail the origin of this pressure and its relation to gas temperature and density. The mathematical expression that we are about to derive will be quite useful in the following for the analysis of cosmic effects.

If a particle of mass m moves with a velocity v, it is said to possess a linear momentum mv. The momentum of a system of particles is a conservative quantity: if no external forces act on the system, momentum is neither created nor destroyed. The conservation of the total linear momentum is one of the fundamental laws of physics. Let N particles hit each cm^2 of the wall each second and suppose that all these particles move with a velocity v perpendicular to the wall. If the incident particles remain stuck to the wall, they transfer a momentum of Nmv per second to the fixed wall. If, however, the particles are reflected from the wall as elastic spheres, they reverse the direction of their velocity. The total change in particle velocity in this case is $v-(-v) = 2v$, and the momentum transferred by the reflected particles each second per cm^2 of the wall surface is $2Nmv$.

According to Newton's law, the rate of change of momentum is equal to the force acting on the object. In this case, the acting force is the force which restrains the wall from moving: it is precisely because of this restraining force that the wall remains fixed and the particles are reflected. This force is numerically equal to the gas pressure, as we have said before, so that the gas pressure is $2Nmv$.

Let the gas be enclosed in a cube, where the particles for some reason can move only at right angles to the cube walls. All the particles have the same velocity. In this case, one sixth of the total number of particles travels toward each of the cube walls. Each second only those particles hit the obstacle which were at a distance not greater than v from the wall a second earlier. Farther particles simply will not make it to the wall. Each square centimeter of each wall therefore intercepts each second one sixth of the total number of particles in a gas column with a base of $1\,cm^2$ and a height of v. If the number of particles in $1\,cm^3$ of the gas is n, this column, having a volume of $v\,cm^3$, will contain nv particles, and the number of particles hitting the wall is $N = nv/6$. The pressure on the wall is therefore $nmv^2/3$.

Our simplifying assumptions concerning the motion of the gas molecules naturally do not correspond to the true state of things. Indeed, the gas particles move at random with various velocities at different angles to the wall. Nevertheless, exact treatment yields the same result, $nmv^2/3$, where v^2 is to be regarded as the mean square velocity of the gas particles. Since by (1) $m\bar{v}^2 = 3kT$, the pressure P is expressed by the relation

$$P = nkT. \tag{2}$$

Equation (2) is simply another form of the Clapeyron law taught in high-school physics. Indeed, the number of particles n in 1 cm^3 is equal to the number of particles in one gram-molecule A (Avogadro's number) divided by the volume V. Inserting for n in (2) the ratio $\frac{A}{V}$ and seeing that $Ak = 6 \cdot 10^{23} \cdot 1.38 \cdot 10^{-16} = 8.3 \cdot 10^7$ erg/deg · mole is in fact the universal gas constant R, we obtain the Clapeyron law of ideal gases: $PV = RT$.

The wall of the gas volume is assumed "ideally reflecting." This implies that the reflected particles move with the same speed as before the collision (although their direction has changed). If, however, the speed changes in collision, i.e., the reflection from the wall is not perfect, the kinetic energy of the particle is also changed. Let the velocity of a particle moving toward the wall be v_1, and the velocity after the collision v_2. If v_2 is less than v_1, the particle has lost a kinetic energy equal to the difference $\frac{mv_1^2}{2} - \frac{mv_2^2}{2}$. This amount of energy is redistributed between the atoms in the wall, making them move faster (the atoms in a solid are in a state of constant vibration). The wall temperature thus increases: the gas lost part of its energy to the wall, heating it up. If the wall temperature is higher than the temperature of the gas, the incident particles gain energy on collision with the wall atoms. As a result the reflected particles move faster than before the collision. A hot solid in contact with colder gas thus loses part of its energy heating the gas.

Let us now return to the problem of explosions. After the source energy — whether chemical, nuclear, or other — has been converted to heat, the pressure of the gas produced by the explosion (or contained from the start in the exploding system) increases drastically. The resulting high pressure propels the nearby object and the gas therefore does a certain work. The explosion energy is thus converted into mechanical (kinetic) energy.

This principle is utilized in internal combustion engines, where the pistons are moved by the gas pressure from a series of explosions in the fuel mixture. Only part of the explosion energy is used for effective propulsion: part is lost in heating the engine.

If the explosion occurs in a solid and the pressure is sufficiently high, the cohesion forces between the particles may not be strong enough to hold the fragments together and the solid breaks up. The detached fragments carry off a certain kinetic energy. A shell explodes in this way, so that eventually the heat is converted into mechanical energy.

The conversion of thermal energy into the kinetic energy of liquids and gases is a much more complex effect. In liquid and gaseous media the motion propagates in the form of a shock wave. These shock waves shatter the glass windows in houses far from the explosion site. We will now explain how a shock wave forms in a gas (shock waves in liquids are not treated, as this case does not arise in cosmic explosions).

A shock wave is a special form of motion in a gas. In daily life we encounter different modes of motion in gases. In particular, propagation of sound involves oscillatory movement of gas. The source produces periodic compression of the gas. The pressure in the compressed layer increases and the gas tends to expand, compressing in its turn the next adjoining layer. A succession of rapidly alternating compressions and rarefactions constitutes a sound wave. Under normal conditions the velocity of propagation of this wave is about 330 m/sec.

The variations of air density in these compressions and rarefactions is very small. One therefore regards the acoustic vibrations in a gas as very weak disturbance of its normal state. The propagation velocity of these disturbances is determined by the molecular motion. After all, the compressed gas layer expands because the constituent molecules move apart, and the next adjoining layer is compressed only when it "senses" the pressure of the molecules from the expanding layer. This explains why the root mean square velocity of a nitrogen molecule that we have calculated above is of the same order of magnitude as the propagation velocity of a sound wave.* Therefore, the higher the gas temperature, the faster is the propagation of acoustic vibrations.

An explosion generally involves a strong disturbance in the state of gas, and this disturbance is regarded as the shock wave. The formation of a shock wave is associated with abrupt and large compression of the gas. This process can be conveniently visualized in terms of a piston which is pushed into a very long pipe filled with gas. Let the piston move so that its velocity increases in discrete jumps (Figure 1): this provides a certain approximation to the general continuously accelerated motion of the piston.

The gas near the piston face is the first to be compressed. Since the initial velocity of the piston is small, the disturbance in the gas is weak and it propagates with the velocity of sound. The disturbance does not encompass the entire gas volume in the tube, and we therefore distinguish between the region of the disturbance (the compression wave) and the undisturbed region.

FIGURE 1. Schematic variation of piston velocity.

Because of compression the gas temperature increases. (This is a well-known effect observed in daily practice.) The work done against the gas when the piston is moved into the pipe is only partly converted into the kinetic energy of the gas. The gas particles collide with the moving piston and are reflected with an increased kinetic energy. The piston thus transfers to the gas molecules some of its kinetic energy, and this energy is then redistributed between other molecules in molecular collisions. The mean particle energy, i.e., the gas temperature, therefore increases. Since the gas molecules have a preferential velocity component in the direction of piston motion (after all they are propelled by the piston to a certain extent), the gas in the compression wave moves as one whole in the same direction with the piston.

When the piston velocity increases in a jump, another disturbance propagates through the gas. Since the gas has been heated by the previous compression, the disturbance now propagates somewhat faster and it catches up with the first compression wave. However, it cannot overtake

* The fit between the two figures is not exact, since air compression is produced not only by molecules moving in the direction of sound propagation, but also by molecules which move at a certain angle, and whose velocity in this direction is smaller.

this compression wave, since the velocity of all sound waves in the undisturbed region (in front of the advancing compression wave) is the same. Having caught up with the first wave, the second wave increases the discontinuity in pressure and gas density across the moving disturbance. Subsequent disturbances produced by further acceleration of the piston continue building up the pressure and density discontinuity between the compression wave and the undisturbed gas.

In ordinary gas flow, the temperature and the density are continuous functions of position, smoothly varying from one point to another. The situation, however, is entirely different in a gas with a shock wave. When the piston velocity is sufficiently high, the gas state is seen to change abruptly and discontinuously (Figure 2). The surface of this discontinuity constitutes the front of the shock wave. The strength of the wave is determined by the pressure discontinuity at the front. The higher the pressure behind the wave front compared to the pressure in the undisturbed gas, the stronger is the shock wave.

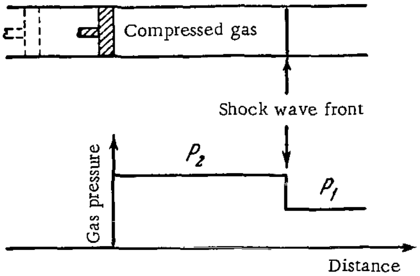

FIGURE 2. Pressure distribution in a gas with a shock wave. The dashed line marks the initial position of the piston.

In practice the shock wave front is a very thin gas layer where the gas state functions — density, temperature, and pressure — change steeply. The wave front moves in the same direction as the piston (though faster than the piston), encompassing ever new gas volumes. The gas is compressed and heated, at the same time acquiring a certain net velocity in the direction of wave propagation.

In explosions, the hot high-pressure expanding gas acts as a piston. If the gas expands isotropically (i.e., uniformly in all directions), the shock wave is spherical, with a center at the explosion focus. After the explosion products have expanded and cooled down, the pressure in the wave falls. The wave is weakened, as it steadily loses energy in heating and accelerating additional masses of gas, and no new energy sources are available.

The mass entrained by a spherical shock wave increases as it moves away from the explosion focus. At a distance R from the focus, the mass is proportional to R^2, i.e., the surface area of the sphere. The strength of a spherical wave therefore increases much faster with distance than, say, the intensity of the plane wave generated by a piston in a pipe.

In the final analysis, the shock wave absorbs not only the thermal energy released in the explosion, but other forms of explosive energy as well. For example, the radiant energy generated in a nuclear explosion briefly heats up the air in the immediate neighborhood of the explosion focus. The pressure increases and the expansion of the hot air produces a shock wave, which propagates over tremendous distances.

The main effects of a strong shock wave are primarily to propel large masses of gas heating them to a point where the gas becomes luminous.

These effects are of the greatest importance in the study of cosmic explosions, since the motion and emission of celestial bodies provides an indication of the strength of cosmic explosions and other characteristics.

In conclusion of this section, before passing to a description of cosmic explosions proper, we would like to dwell on some quantitative parameters of explosions. Explosions naturally differ in the quantity of energy released. A useful unit is the energy released in the explosion of 1 kg of trinitrotoluene (TNT) — a common high explosive. It is equal to 10^6 calories. One calorie is equivalent to $4.18 \cdot 10^7$ erg (CGSE energy units) or 4.18 joule (MKS energy units).* The explosion of 1 kg of TNT thus releases $4.2 \cdot 10^{13}$ erg or $4.2 \cdot 10^6$ joule of energy. In terms of energy output, a megaton hydrogen bomb is equivalent to a million tons or 10^9 kg TNT. The explosion of a megaton bomb therefore releases $4.2 \cdot 10^{22}$ erg of energy.

The effect of the explosion on the surrounding medium depends not only on the absolute quantity of energy released, but also on how fast it is released. Dividing the explosion energy by the time in which it is released, we obtain a quantity known as the explosion power. For example, if a megaton bomb explodes in one millionth of a second, the explosion power is $4.2 \cdot 10^{22}/10^{-6} = 4.2 \cdot 10^{28}$ erg/sec or $4.2 \cdot 10^{21}$ joule/sec. A unit of power corresponding to 1 joule per sec is known as the watt, and 1000 watts are 1 kilowatt. The power of a megaton bomb is thus $4.2 \cdot 10^{18}$ kW, whereas the output of the largest hydroelectric power stations does not exceed a few million kilowatts. To produce the energy equal to that released by a single megaton explosion, a power station with an output of 4 million kilowatt should work continuously for more than 10 days. The explosion power is particularly significant for assessing the damage that the explosion causes.

§ 2. METHODS OF STUDY OF COSMIC EXPLOSIONS

Terrestrial explosions generally can be probed without much difficulty. The energy causing the explosion can readily be established both for artificial and natural explosions. The main attention is usually focused on the effects accompanying the explosion, such as shock waves, acoustic effects, light emission. These effects are directly recorded with various instruments.

The situation is not as simple for cosmic explosions: it is not always possible to decide what is the exact energy source of the explosion. This problem can be solved only after a careful study of all the observable effects produced by the explosion.

The difficulties associated with the study of cosmic explosions, as of any other astronomical phenomenon, are mainly due to the great distance of the object of study from the observer. Even the distance of the nearest celestial body, the Moon, is more than 9 times the length of the Earth's equator. The Sun is 400 times as distant from the Earth as the Moon, namely at a distance of 150 million kilometers. And yet, even this figure is miniscule compared to the distances of the nearest stars.

* In what follows we use the CGSE system, and all the values can be converted to the MKS system if desired using this relation.

Any measurement instrument used to probe the explosion responds effectively only when it receives a fraction of the energy released in the explosion. An earthbound observer receives a vanishingly small fraction of the energy released in a distant explosion in the far reaches of the Universe, because the energy moving away from the source fill an ever increasing volume. The energy flux F is defined as the quantity of energy crossing each second a 1 cm^2 surface area at right angles to the direction of energy propagation. We will now establish how F varies with distance from the explosion site. If the carriers of energy are the particles accelerated by the explosion, which are assumed to move with constant velocity, the energy through any sphere centered at the explosion focus is naturally constant. On the other hand, the energy per cm^2 is inversely proportional to the area of the sphere, i.e., to the square of its radius. The energy flux at any point is therefore inversely proportional to the square of the distance from the explosion site, i.e., for the fluxes F_1 and F_2 at distances r_1 and r_2 from the focus we have

$$\frac{F_1}{F_2} = \frac{r_2^2}{r_1^2}. \tag{3}$$

This expression is valid irrespective of the particular form of energy involved, as long as it propagates uniformly in all directions (or inside a cone with its apex at the explosion focus). In particular, for light energy equation (3) gives the elementary law of decrease in luminous intensity with the distance from the light source. If on its way to the observer the energy is converted to other forms, e.g., mechanical energy is converted to heat, the energy flux diminishes faster than prescribed by equation (3).

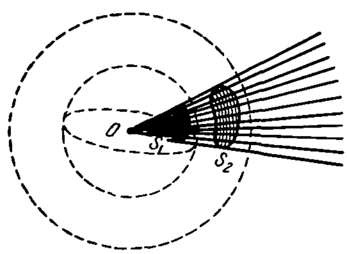

FIGURE 3. Propagation of energy from the focus of an explosion. The energy flux is inversely proportional to the surface area of a sphere centered at the focus.

Earthbound instruments are incapable of picking up the mechanical energy of even the most powerful stellar explosions. Shock waves propagating in the interstellar gas are damped over comparatively small distances from the explosion focus. Moreover, the shock wave velocities are hundreds and thousands of times less than the propagation velocity of radiation. It takes several years for the light of the nearest stars to reach the Earth. The mechanical energy, even if it were to reach the Earth in measurable amounts, would be delayed to such an extent that we would be unable to assign it to any particular explosion.

Electromagnetic radiation (light being one of its forms) is actually the only form of energy emitted by stellar explosions which can be detected directly on the Earth without much difficulty. In principle, we can also detect particles which move with velocities near the velocity of light, such as the neutrino, but unfortunately the modern detectors of high-energy particles are not sensitive enough for "seeing" the exceedingly weak particle fluxes which should reach the Earth from the exploding stars. Moreover, charged particles (electrons, protons) do not move in straight lines in space because of various interactions, and their origin can hardly be assigned to some particular stellar explosion.

The only practicable source of information about cosmic explosions is thus still the electromagnetic radiation. These explosions are therefore studied by the conventional methods of astronomy, and the object of the present section is to describe how the radiation reaching the earthbound observer from celestial objects can be interpreted to provide useful information.

Electromagnetic radiation is generated by nonuniform (accelerated) motion of electrically charged particles. Any charged particle sets up an electric field in the nearby space. The electric field is in fact that region of space where the source particle measurably interacts with other charged particles. Moving charges also produce magnetic fields. This is a well-known experimental fact. We all know that a conductor through which electric current is flowing exerts a certain force on a permanent magnet, which indicates that the current conductor generated a magnetic field. Electric current is made up of moving charged particles, so that any moving charged particle generates two fields — electric and magnetic.* These fields can be treated as a single electromagnetic field.

When the charged particle changes its velocity (either its magnitude or direction), the electromagnetic field also changes. Any change in the electromagnetic field propagates with a velocity of 300,000 km/sec (this velocity is generally designated by the letter c). These field changes constitute what is known as electromagnetic waves. For example, an electron escaping from an atom moves in a straight line until it meets a positively charged proton, which deflects it from its course. The attraction that the electron experiences in the field of a proton now makes it move in a hyperbolic trajectory. The kinetic energy of the electron decreases as it moves: part of its energy is converted to electromagnetic field energy and dissipates in space. This process is known as emission of radiation in free-free transitions, since a free electron remains free in space: it did not attach itself to the proton but only changed the direction of its motion.

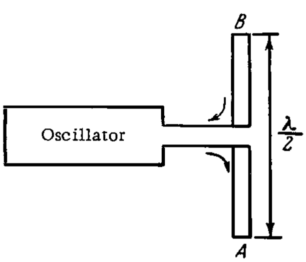

FIGURE 4. A dipole source. The oscillator periodically reverses the direction in which the electrons in the dipole are driven.

An oscillating electric charge, which moves with a certain period, also generates a periodic field. The propagating field disturbances in this case are electromagnetic oscillations, which constitute a highly important case of electromagnetic waves: any field variation, and in particular the radiation emitted in free-free transitions, can be represented as a sum of oscillations.

A very simple generator of electromagnetic oscillations — called a dipole oscillator — can be made of two lengths of conductor connected to an oscillator (Figure 4).

* In fact the field of any charged particle fills the entire space, but the field strength falls off rapidly and the field is measurable only fairly near the source.

The oscillator drives the electrons in the conductors alternately in opposite directions. The energy of the electromagnetic waves emitted by this source is the result of the work done by the driving force which periodically changes the direction of the electrons in the conductor.

The period of the oscillations is equal to the period of the force acting on the charges. The oscillations, however, are better characterized in terms of their frequency, which is equal to the number of oscillations in 1 sec. To increase the oscillation frequency, i.e., to make the electrons reverse their direction more often, a larger work has to be done in the same time. Therefore, the radiation power increases with increasing frequency of electromagnetic oscillations, and this conclusion is equally applicable to the oscillation of a large assembly of charged particles and of a single electron.

Alongside with the oscillation frequency ν, we will often use the wavelength λ. The wavelength is related to frequency by the simple equality

$$\lambda = \frac{c}{\nu}. \tag{4}$$

The meaning of equation (4) is self-evident. Indeed, in one second the electromagnetic wave covers a distance equal to c, and all the waves emitted during that second are accommodated within that distance. The number of waves is ν, and therefore the distance between the corresponding points of successive waves, between their crests say, is $\frac{c}{\nu}$.

The wavelength of electromagnetic radiation is determined by the particular oscillator used. The wavelengths are not restricted from above, i.e., arbitrarily long electromagnetic waves are possible. The properties of radiation depend on λ. We therefore distinguish between different kinds of electromagnetic radiation according to the range of wavelengths (or the frequency band). Thus radiation with wavelengths from a few thousand meters to a few millimeters is classified as radio waves. The radio waves in their turn are divided into long, medium, short, meter, centimeter, and millimeter waves.

The visible radiation covers a fairly narrow range of wavelengths, from $\lambda = 4 \cdot 10^{-5}$ cm to $\lambda = 7 \cdot 10^{-5}$ cm. This is the so-called optical or visible spectrum. Not more than some thirty—forty years ago, the astronomy was essentially confined to this range of wavelength. At $\lambda = 4 \cdot 10^{-5}$ cm the radiation is sensed as having violet color, and at $\lambda = (6.5-7) \cdot 10^{-5}$ cm the radiation is red. Wavelengths over $7 \cdot 10^{-5}$ cm correspond to the so-called infrared radiation. The infrared spectrum at its long-wave end joins the millimeter radio waves.

The division of the electromagnetic spectrum into different frequency bands is not rigid. It is mainly associated with the fact that different radiation detectors respond essentially to different wavelengths. Thus, radio receivers are entirely insensitive to optical radiation, whereas the human eye does not "see" radio waves. The transmission of radiation through the Earth's atmosphere is also different at different wavelengths. The atoms and molecules of atmospheric gases scatter and absorb electromagnetic radiation, but the scattering and absorption processes essentially depend on the frequency of the electromagnetic oscillations. For example, optical radiation is readily scattered in clouds, but the scattering is not large in a clear atmosphere. Radio waves, on the other hand, are absorbed to a much greater extent by the water molecules in the atmosphere, and this absorption is significant even in a clear atmosphere without any clouds. At meter and centimeter wavelengths the Earth's atmosphere is fairly transparent. Longer wavelengths, on the other hand, are absorbed and reflected by the outermost layers of the atmosphere. This effect is discussed in some detail in § 5.

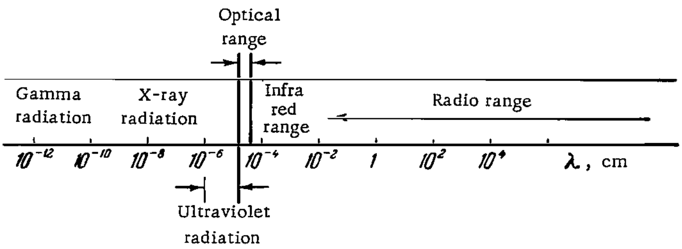

FIGURE 5. The electromagnetic spectrum.

Thus the range of wavelengths automatically determines the particular detector to be used and the feasibility of detection at the ground, under the atmosphere. The lower atmospheric layers are perfectly opaque to radiation with wavelengths between $3 \cdot 10^{-5}$ and 10^{-6} cm. The radiation of celestial bodies in this range (ultraviolet radiation) can be detected only above the lowermost atmospheric layers; the measurements in the ultraviolet are therefore carried out with the aid of suitable equipment from rockets and artificial Earth satellites. Yet even in the upper atmosphere only the ultraviolet radiation of the Sun and the planets can be detected at wavelengths between $\lambda = 9 \cdot 10^{-6}$ and 10^{-6} cm: the radiation from stars at these wavelengths does not reach the Earth, since the tenuous gas filling the interstellar space stops it effectively.

The part of the electromagnetic spectrum corresponding to λ from 10^{-6} cm to 10^{-8} cm is known as the X-ray region, and even shorter wavelengths correspond to gamma rays. These forms of radiation do not penetrate through the Earth's atmosphere and like the ultraviolet radiation they are observed in the upper atmosphere.

The wave nature of electromagnetic radiation is confirmed by a wide range of experiments. Other experimental findings, however, seem to indicate that radiation sometimes behaves as a stream of particles. There is no contradiction in these results, and they are a reflection of the complex features of the phenomenon. Electromagnetic radiation cannot be fully described in terms of particles only or wave motion only. There is inherent duality in radiation, which reveals the properties of both particles and waves.

In the following we will often treat radiation as a stream of particles. These particles are known as radiation quanta or photons. The photons always move with the velocity of light c, and each photon has an energy $h\nu$, where h is Planck's constant, equal to $6.62 \cdot 10^{-27}$ erg · sec in the CGSE system of units. For a given energy flux, the number of photons decreases as the frequency increases. Therefore the particulate properties of radiation are the most prominent at low wavelengths and high frequencies. Radiation of radio frequencies is conveniently regarded as waves. Ultraviolet, X-ray and especially gamma-ray radiation mainly behave as particles. This point is of great significance in designing suitable radiation detectors.

Although ultraviolet and X-ray detectors are constantly improving, virtually the entire information on celestial bodies and cosmic explosions is obtained in the optical and the radio spectrum.* It is these forms of radiation that are considered in what follows. Optical radiation is recorded with photographic plates and photocells, which supplement the human eye as detectors. The main shortcoming of photographic plates is their low efficiency in the optical region. Of the thousand photons hitting the plate, only one or two leave a recordable trace. A photocell is much more effective in this respect. Photocells convert the radiation energy into kinetic energy of electrons, and modern devices of this kind used in astronomy effectively record one or two photons of each ten photons hitting the cell.

The human eye, though an exceptionally sensitive radiation detector, is deficient as a measuring instrument: the results of low-flux observations cannot be recorded rapidly and automatically. Eye observations are therefore seldom made these days. Other optical detectors are also available, but we will not discuss them here.

Telescopes are used to step up the radiation energy reaching the detector. A reflecting telescope has a parabolic mirror which focuses all the incident radiation at one point. The quantity of light collected by a telescope relates to the quantity directly reaching the human eye as the mirror area to the eye pupil area. The largest of the modern optical telescopes has a 200-in. mirror, i.e., a mirror with a diameter of 5 meters. The mirror and pupil areas relate as the squares of the respective radii. Since the radius of the pupil is about 0.5 cm, the telescope mirror collects $(500)^2/(0.5)^2 = 10^6$ more light than the human eye does.

Radio telescopes are used to enhance the received energy in the radio frequency range. The radio waves are collected by a metallic antenna — a parabolic dish or a plane reflector. The diameters of parabolic antennas reach tens of meters, whereas the plane antennas measure hundreds of meters. Since the area of a radio telescope is much greater than the area of an optical reflector, they are capable of detecting exceedingly weak sources of radiation, which are at present inaccessible to optical observations. The radio energy collected by the antenna is transmitted through a coaxial cable to a receiver, where the signals are amplified, modulated, and then recorded in a suitable form.

What information does electromagnetic radiation provide? The observations naturally determine the position of the radiation source in the sky, the shape of the source, and if its distance from the Earth is known also the size and the energy output of the source. This, however, is generally insufficient to establish the physical state of the celestial body or the reasons behind the explosion. Much more extensive information on

* The only exception in this respect is the Sun, as its ultraviolet and X-ray radiation are now successfully measured from artificial satellites and rockets.

the physical state of celestial bodies can be obtained if we measure the radiation in narrow wavelength regions, rather than the total integrated radiation of the source. To understand what is involved, let us study in some detail the physical processes responsible for the emission of radiation in various objects. Since the great majority of celestial bodies are made up of high-temperature gas, we will only concentrate on emission processes of hot gases.

A gas like any other physical system is made up of atoms, which are assemblages of various charged particles. The positively charged atomic nucleus is surrounded by negatively charged electrons. If the nucleus charge is equal to the total charge of the electrons, the atoms are neutral; otherwise we have an ion with a net electric charge. The simplest atom is that of hydrogen, which has only one electron. The nucleus of the hydrogen atom is called a proton. The next atom in the order of increasing complexity, that of helium, has a nucleus with four times the mass of the hydrogen nucleus: it consists of two protons and two neutral particles called neutrons, whose mass is close to that of the proton. There are two electrons in the neutral helium atom.

An electron moving in the field of a nucleus cannot have just any energy. For each atom there is a certain set of energy values that the electron may assume, the so-called energy states of the atom.

The nucleus and the electrons experience mutual attraction, as oppositely charged particles always do. This attraction is balanced by another force, associated with the motion of the electron around the nucleus.* To move an electron away from the nucleus to a farther orbit of a larger radius, work should be done against the electrostatic force attracting it to the nucleus. If, conversely, the electron moves closer to the nucleus, a certain energy is realesed in the atom. As the electron may occur only in certain discrete energy levels in the atom, it absorbs and emits energy in discrete portions, or quanta, which are equal to the difference in energy between the initial and the final state. To each state of the atom corresponds a certain definite value of electron energy, and the transition of the atom from one energy state to another corresponds to the transition of an electron between corresponding levels.

If no external fields and forces act on the atom, the electron will move in the orbit nearest to the nucleus. This state of minimum energy is called the ground state of the atom. An atom in its ground state does not emit any energy: after all the electron cannot drop to any lower energy level. Note that as we have mentioned before, any charged particle describing a nonlinear trajectory emits electromagnetic radiation, so that the electron moving in a circular or elliptical orbit is nevertheless expected to emit. This is not so, however: the physical laws governing the microcosmos — the world of atoms and electrons — are different from the laws of the macrocosmos. Atomic systems are governed by the laws of quantum mechanics, which state that an electron in its stationary orbit does not emit any radiation.

The gas molecules collide and exchange kinetic energy. These collisions, however, are not always like the collisions of elastic spheres. If an atom collides with a particle whose kinetic energy is greater than

* This is a highly inaccurate description of the forces acting in an atom, but it is adequate for explaining the processes of gas emission.

the difference in energy between the ground state and one of the allowed states, the atom may be excited from the ground state to one of the higher-lying allowed states. This is how excitation of atoms occurs. In the case under consideration, the atoms are excited by the thermal energy of the gas.

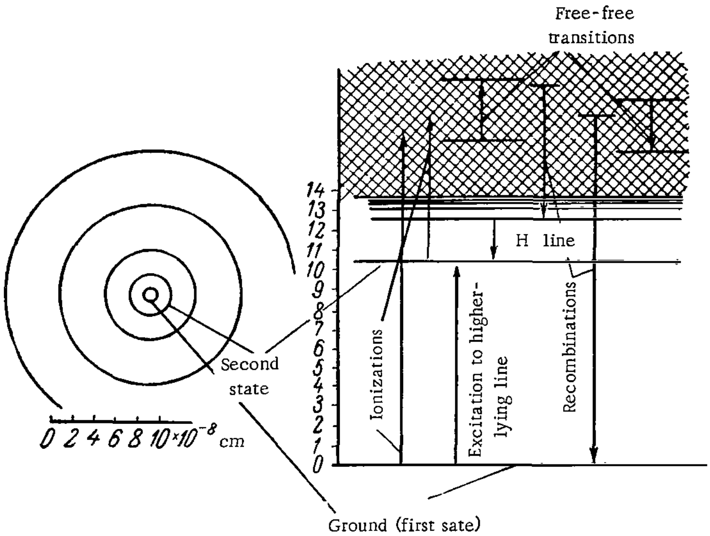

FIGURE 6. A hydrogen atom and its energy level. Left: circular electron orbits (drawn to scale); right: a diagram of energy levels, showing various electron transitions. The crosshatched region corresponds to a free electron (hyperbolic orbits). The state energies in electron-volts ($1\,\mathrm{eV} = 1.60 \cdot 10^{-12}$ erg) are marked on the left; the energy is reckoned from the ground state, which is assumed at zero energy.

An atom remains excited generally for a very short time, of the order of 10^{-9} of a second, and then it reverts to the ground state, either directly or in successive transitions through an intermediate state. In these de-excitation transitions, the atom emits electromagnetic radiation, in the form of one or several photons. The total energy of the emitted photons is equal to the kinetic energy absorbed in the excitation of the atom. In the final account, part of the thermal energy of the gas has been converted into radiant energy.

If the kinetic energy of a colliding particle is sufficiently high, an electron can be moved to a large distance from the nucleus, and the atom loses it irretrievably. This process of ionization by collision produces an ion and a free electron, not bound to any atom. The free electrons are now one of the gas constituents, and they can exchange energy with other particles. Excitation and ionization in general can be caused by collisions with any particle, as long as its energy is sufficiently high. The electrons, however, are particularly effective in this respect.

Reverse processes also take place in the gas, i.e., radiant energy is converted to heat. This occurs, for instance, when an atom encounters a photon whose energy is higher than that required for ionization. In this case an electron is detached from the atom, and the excess energy of the photon imparts a certain kinetic energy to the electron; this kinetic energy is then redistributed as heat between all other particles. A different process of energy conversion occurs when a photon only excites an atom, without ionizing it. If during the time that the atom is in a state of excitation it collides with another particle, its excitation energy will be transmitted to that particle and the atom will drop to the ground state without emission of radiation. Since a photon has truly vanished in the process, we say that this is a case of true absorption of radiation, i.e., radiant energy has been converted to thermal energy of the gas.

Atoms can be made to emit only if some energy has first been pumped into them. Atoms are excited not only by the gas particles, but also by extraneous particles, provided they are sufficiently energetic, and by incident electromagnetic radiation — a flux of photons. Excited atoms will then emit photons whose energy is precisely equal to the energy difference between two allowed states of the atom, so that the frequency of the emitted radiation is naturally restricted.

FIGURE 7. Hydrogen emission spectrum.

The quantity of emitted energy depends on the wavelength. This dependence is known as the emission spectrum of the body. The emission spectrum of a gas without any ionized atoms shows radiation concentrated in very narrow frequency intervals. These emission peaks are known as the emission lines of the gas. The width of the corresponding frequency interval is known as the line width. Each line corresponds to a transition between two energy levels of an electron in an atom, i.e., a photon corresponding to the line frequency is emitted when an electron jumps from a high-energy level to a low-energy one. The line width, although small, is nevertheless finite, since the energy of the atom in an excited state is not exactly defined. The quantum-mechanical rules which govern the microcosmos state that the energy can take values from a narrow, but nevertheless finite, interval, and the energy level is said to be "blurred," not sharp.

The emission spectrum of a gas which contains ions and free electrons (this gas is called a plasma) is different from the spectrum of non-ionized gas. The plasma electrons combined with ions to form neutral atoms — this process is known as atomic recombination. The total energy of an atom formed by a free electron and an ion is less than the sum of the energies these particles had before recombination: after all, energy is required to split the atom into two separate particles. The excess energy escapes in the form of radiation from the electron — ion system during recombination. Ions can trap electrons with any kinetic energy, and the excess energy is therefore different in each case. Recombining plasma will therefore emit photons in a wide range of frequencies, and not in narrow lines. This constitutes what is known as a continuous spectrum.

The emission of a plasma in free-free transitions is also continuous. The change in electron energy in such a transition depends on the conditions of collision with protons. Therefore different collisions emit photons of different energies, and a continuous spectrum is obtained.

Strong interactions between atoms are another source of a continuous emission spectrum in high-density media, whether solid or gaseous. These interactions shift the atomic energy states and cause overlapping of spectral lines.

If the continuous radiation emitted by a body traverses some tenuous gas on its way to the observer, the appearance of the spectrum is markedly complicated. The body emits photons of any frequency, but the gas atoms are excited only by those photons whose frequencies correspond to one of the spectral lines. It is only in this case that the photon energy is converted to the excitation energy of the atom. The reverse transition of an excited atom to the ground state emits a photon of the same frequency, but it escapes in a quite arbitrary direction, and the probability is very low that it will be emitted in the primary direction.* A photon originally emitted in the direction of the observer will therefore be scattered in this process, without reaching the observer.

There is still another possibility: the excited atom drops to the ground state through several intermediate transitions. The electron jumps down from one level to another until it returns to the ground state. The original exciting photon thus disappears, and several photons of lower energy are emitted.

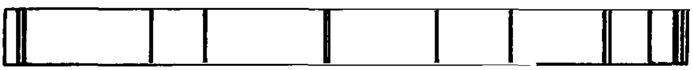

FIGURE 8. Spectrum with absorption lines (part of the solar spectrum).

Finally, an excited atom may transmit its excitation energy to some other particles it collides with. The photon again disappears in this case, as its energy is converted to the thermal energy of the gas. All these processes reduce the number of photons reaching the observer at a given line frequency compared to the number of photons of other frequencies. These photons do occasionally disappear, e.g., in ionization of atoms, but to a much smaller extent. The ability of an atom to capture an ionizing photon is immeasurably less than its ability to become excited. For each absorbed continuous spectrum photon, there are tens of thousands of absorbed photons in nearby lines. A dark narrow line is therefore observed against a bright, virtually unattenuated background at the frequency corresponding to the spectral line of the gas. These dark lines are called absorption spectrum lines, and they constitute a highly prominent feature

* If the number of photons at a given line frequency is high, the interaction with the excited atom can be such that the scattered photon will be emitted in the same direction as the incident photon (so-called stimulated radiation). In celestial bodies this effect is significant in a number of cases, especially at radio frequencies.

in the emission spectra of stars. The absorption lines in the spectrum of the Sun were discovered as early as the beginning of the 19th century; they are known as Fraunhofer lines, after their discoverer.

The interaction of radiation with matter, as we see, should produce a number of characteristic features in the observed spectra of celestial bodies, which will enable us to determine the physical properties of these objects. To obtain a spectrum, the light collected by a telescope from a star is generally directed into a special instrument, known as a spectrograph. The commonest prism spectrograph uses the property of the prism to refract light of different wavelengths to a different extent. The radiation collected by the telescope is a mixture of various wavelengths. This mixed light is directed as a narrow beam to a prism. The photographic plate mounted behind the prism sees an elongated strip because of the unequal refraction of light.* Each wavelength corresponds to a certain position in this strip, and one can thus study the distribution of radiation over the different frequencies. Radiation in fairly narrow spectral intervals can also be observed without any spectrograph, by using filters which transmit light at a certain wavelength only, while blocking out the rest of wavelengths.

The spectrum makes it possible to determine the energy emitted by the body at various wavelengths. From these data, especially for the spectral lines, we can find the temperature of the emitting gas, its pressure and chemical composition.

Determination of the chemical composition is based on the following fact: the atoms and ions of each element have their own system of spectral lines associated with the particular set of energy levels of the given atoms. The spectra of atoms and ions of various elements have been studied in laboratory and calculated theoretically. Therefore, observing a certain sequence of spectral lines in the spectrum we can identify the element to which it belongs. This naturally proves the occurrence of that element in the celestial body being studied. The spectra of celestial bodies generally contain a great multitude of lines and it is difficult to decide which of the lines belong to what element, especially since only some of the lines of each element fall in the optical spectrum.

When the qualitative composition of the emitting body has been determined, the relative content of the atoms of various elements is found by a comparison of the corresponding spectral lines. As a rule, the greater the number of a particular atomic species in the emitting (or absorbing) gas layer, the more prominent are its lines against the continuous emission background. The principle of these observations is fairly simple, but their application in practice requires a great amount of work.

The relative intensity of the spectral lines of any given element depends on the temperature of the gas. As the kinetic energy and the particle velocity increases (i.e., as the temperature of the gas increases), atomic excitations become more frequent. The energy quanta received by colliding atoms become larger, and the electrons in atoms are moved to progressively higher energy levels. The proportion of photons associated with transitions from higher-lying levels therefore increases, and the form of the spectrum changes accordingly.

* A prism is sometimes used to intercept the light at the entrance aperture of the telescope. The spectra obtained in this way are not particularly suitable for detailed investigations.

As the temperature of the gas increases, the ionized atoms become more numerous, since the ionizing power of particles is higher. Therefore the lines of ions develop and intensify at high temperatures. An ever increasing energy is required to detach each successive electron from the atom, and therefore the appearance of lines of doubly or multiply ionized atoms in the gas spectrum is a sign of sufficiently high kinetic energies or temperatures.

Comparison of observation results with laboratory findings and theoretical calculations of atomic structure shows what proportion of atoms occurs in each particular energy state and what is the content of ionized atoms relative to the neutral atoms. The temperature of the emitting body is inferred from these measurements.

The content of ions relative to the concentration of neutral atoms of the same element, the so-called degree of ionization, depends on the gas density, as well as temperature. This fact is used in density determination from the lines of neutral atoms and ions of the same element. At higher temperatures, the lines of singly and doubly ionized atoms are compared.

The information obtainable from spectroscopic observations of celestial bodies is not limited to the above parameters. Spectral lines also provide an indication of magnetic fields in the gas and make it possible to determine the strength of these fields. These measurements became very popular in recent years, as magnetic fields were found to play an important part in various processes in the Universe, cosmic explosions included.

It was established back in 1896 that in a gas immersed between the poles of a strong magnet each emission line is split into several components. The distance between these components is a function of the magnetic field strength. This splitting, known as Zeeman effect (after the name of its Dutch discoverer), is associated with the influence of the magnetic field on the motion of electric charges. The field thus affects the motion of electrons in the atom. The magnetic field alters the set of allowed energy states: each level is replaced by several closely spaced sublevels, and the energy difference between them is proportional to the field strength. Thus, instead of the one line corresponding to electron transitions from one energy level, we observe several lines, and the distance between the line components is found to be proportional to the field strength.

An external magnetic field not only splits the spectral line, but also imparts a special property to the radiation emitted in each of the components. This radiation is found to be polarized.

Polarization is the most obvious manifestation of the wave nature of electromagnetic radiation. The propagation of electromagnetic waves involves periodic variation of field strength in the direction of propagation. Since field strength is a force, it has both magnitude and direction, and the wave propagates in the direction perpendicular to the directions of the electric and the magnetic fields. If at any point along the wave the electric field strength (field vector) lies in the same plane (Figure 9), the wave is said to be linearly polarized.

Although the radiation emitted by each atom has a certain plane of oscillation, which depends on the orientation of the atom, and its light is therefore polarized, the radiation reaching us from celestial bodies is generally not polarized. The point is that this radiation is produced by a multitude of atoms which are randomly oriented one relative to the other and whose orientation relative to the observer constantly changes. Therefore, no preferred plane of the electric vector can be found in the observed radiation.

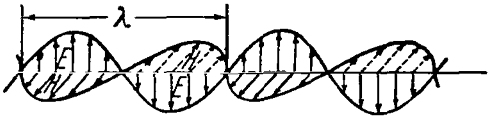

FIGURE 9. Propagation of linearly polarized electromagnetic radiation. The electric vector E lies in the same plane at all points of the wave.

Polarized light can be obtained by natural means from unpolarized radiation. Thus, when electromagnetic radiation is scattered by atoms, the light propagating at right angles to the direction of incidence is seen to be polarized. Polarization of light is detected with the aid of certain crystals which only transmit radiation with the plane of the electric vector parallel to a certain direction, known as the crystal axis. If the quantity of transmitted light varies when the crystal is rotated, the light is seen to be partly polarized, i.e., it is a mixture of linearly polarized and unpolarized light. If at a certain position of the crystal no light is transmitted altogether, the light is fully linearly polarized.

The discovery of linear polarization of the light from some celestial bodies led to a number of important conclusions concerning the nature of cosmic explosions. This point is discussed in a later section in the book. The same also applies to the polarization produced in connection with the Zeeman effect, although in this case the radiation is circularly polarized. The electric field vector in circularly polarized radiation may change its direction, although in a certain way only: the tip of the field vector describes a circle. Depending on whether the vector turns clockwise or counterclockwise (relative to an observer looking against the direction of propagation), we speak of left-hand or right-hand polarization. The line emission split by a magnetic field shows circular polarization when viewed in the direction of the field. The two side components of the split line (it is only these components that are visible in these observations) are oppositely polarized, and the sense of circular polarization depends on whether the field is directed left to right or right to left. Observations in a plane perpendicular to the field shows three linearly polarized components. Observations of line splitting in magnetic fields thus provide an indication of both the field strength and the field direction.

As we have mentioned before, a highly important characteristic is the quantity of energy released in the explosion. The kinetic energy released in a cosmic explosion can be determined from the velocities of the participating objects. Sometimes this is done in direct observations through a telescope. In most cases, however, the velocities are measured from the spectrum. These measurements use the so-called Doppler effect, according to which the frequency of the radiation received from any source depends on the velocity of the source relative to the observer. This dependence is easily derived from elementary calculations.

If the radiation source and the observer are fixed one relative to the other, the radiation frequency received by the observer is no different from the emitted frequency, which we denote by ν_0. The light covers the distance from the source to the observer in a certain time t, and this distance is therefore ct. Now let this distance increase with a velocity v. During the time t, the moving source emits the same number $t\nu_0$ of waves as the fixed source. Previously, however, all these waves were accommodated in a length ct, whereas now, with a moving source, the accommodating distance is $ct + vt$. With a fixed source, the wavelength λ_0 of the radiation received by the observer is $\frac{ct}{t\nu_0}$. If now the source and the observer recede from one another, the wavelength of the received radiation will be $\lambda = \frac{ct + vt}{t\nu_0}$. Therefore λ and λ_0 are related by the equation

$$\lambda = \lambda_0 + \lambda_0 \frac{v}{c}. \tag{5}$$

When the source recedes from the observer, the radiation wavelength increases and the frequency correspondingly decreases. When the source moves toward the observer, the wavelength of the received radiation is shortened. In continuous emission this effect is not noticeable, but it is highly prominent in narrow spectral lines. Depending on the direction of motion the spectral line formed by the moving source shifts toward shorter wavelengths (the violet end of the spectrum) or toward longer wavelengths (the red end). The amount of the shift is $\lambda_0 \frac{v}{c}$.

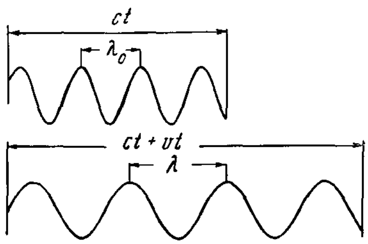

FIGURE 10. Change in wavelength due to source motion.

The normal position of the line in the spectrum, i.e., its wavelength λ_0, is determined experimentally or by calculation. Comparison of the actual line position in the observed spectrum of a celestial body with its normal position gives the line shift $\lambda_0 \frac{v}{c}$ and hence the velocity v of the celestial body in the direction of the observer's line of sight is calculated. These observations are not very exact on the terrestrial scale: the error is never less than 1 km/sec. However, since the celestial bodies move with velocities of tens and hundreds of km/sec, this accuracy is quite sufficient. Unfortunately the Doppler shift method only gives the line-of-sight component of the velocity, and not the total velocity of the object.

All this is also applicable to radio waves. The radio-frequency radiation, however, gives entirely different information on the physical state of celestial bodies compared to observations in the optical spectrum. The point is that most of the line frequencies of the atoms fall in the optical and the ultraviolet spectrum. Observable radio lines are very few, but radio observations nevertheless provide valuable specific information which is not available at other wavelengths. Moreover, radio waves have a most important property: they readily penetrate through the atmosphere, which absorbs optical radiation, and are not attenuated by the cold gas and dust clouds filling the interstellar space, which are opaque in visible light. Our knowledge of the remote parts of the Galaxy is therefore entirely based on information obtained from radio observations.

§ 3. THE STATE OF MATTER IN THE UNIVERSE

A cosmic explosion involves a rapid change in the physical state of the exploding object. Therefore, before proceeding with a description of various explosions, we should concentrate on the different forms of matter and energy encountered in celestial objects. A discussion of energy is deferred to the next section, and here we will briefly describe the structure of celestial bodies, i.e., the state of matter in the part of the Universe accessible to our observations.

Explosions may occur on any celestial body, including the Moon and the planets of the solar system. In this book, however, we will ignore the explosions on planets and satellites, since they are fundamentally related to terrestrial explosions (volcano eruptions, earthquakes, etc), and not to stellar explosion processes. The structure of the solar system therefore is not described here, and the reader is referred to the extensive popular literature which has been published on the subject.

Observations show that the Universe is mainly composed of stars, i.e., the bulk of matter is concentrated in stars. This explains why our main attention is focused on stars and the Sun. Although the Sun is quite an average star, we know a great deal more about its nature and behavior due to its relative proximity to the Earth.

First we should familiarize ourselves with the scale of the stellar worlds, i.e., the typical distances between the stars and the sizes of stars and star systems. These scales are substantially different from the normal terrestrial scales, and special methods are naturally used in distance measurements in the outer space.

Historically the first and until recently the most important method of distance measurements in the solar system and for the nearby stars is the method of the so-called parallactic displacement. As we all know, when an observer moves, the nearby objects appear to be displaced relative to the most distant background objects. The magnitude of this apparent parallactic displacement depends on how far the observer moved and on the distance to the object. For equal displacements of the observer, the parallactic displacement decreases for progressively farther objects.

The distance from the Earth to the other objects in the solar system is negligible compared to the average interstellar distances. When the observer moves over the Earth's surface, the direction to any fixed star does not change. On the other hand, two observers viewing the Moon from different places on the Earth have to look in different directions, and the parallactic displacement can be measured against the background stars.

The distance between the observers and the angular displacement of the Moon (on passing from one observer to the other) is measured without difficulty, and we can then calculate the distance from the Earth to the Moon. On the average it is found to be 380,000 km.* This distance is equal to more than sixty times the Earth's radius. The distance of the Earth from the Sun has been measured in a similar way. It is close to 150,000,000 km, which is 400 times the distance to the Moon.

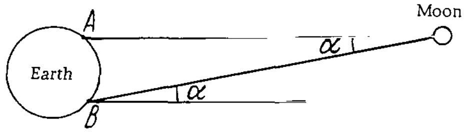

FIGURE 11. How the distance to the Moon is determined. Direct measurements give the angle α and the arc AB on the Earth's surface.

As we have already noted, by travelling between different places on the Earth the observer will not notice any displacement in the apparent position of stars. However, the orbital motion of the Earth around the Sun carries the observer over a distance of 300,000,000 km in half a year (this is the length of the diameter of the Earth's orbit), and this distance is large enough to affect the apparent position of the nearer stars. These parallactic changes are minute, less than one second of arc, but high-accuracy astronomical instruments will measure these angles as long as they are greater than 0.''01. These parallax measurements established that the nearest stars are thousands of times more distant than the Sun. To cross this distance, a photon moving with a speed of 300,000 km/sec will have to travel several years. This is the order of magnitude of the interstellar distances, and new units of length are introduced in astronomy for convenience: these are the light year and the parsec. One light year is equal to $9.5 \cdot 10^{17}$ cm, and one parsec (1 pc) is 3.26 light years. The distance to the nearest star (the star α in the constellation of Centaurus) is 4.3 light years.

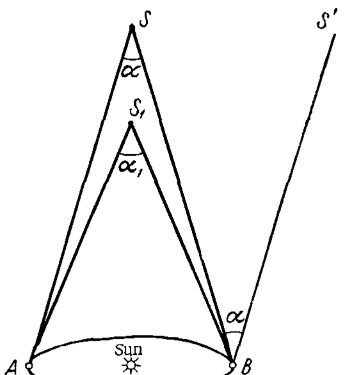

FIGURE 12. The parallactic displacement of stars due to the orbital motion of the Earth around the Sun.

Over these tremendous distances all stars appear as brilliant points of light. Even the most powerful instruments do not resolve the disk of the star. The diameters of all stars, with the exception of the Sun, are therefore measured by indirect techniques, based on temperature determination. The temperatures, as we have mentioned before, are found from the spectrum of the celestial object.

* Since the Moon does not describe a circular orbit, its distance from the Earth is variable.

The spectra of almost all stars contain a multitude of absorption lines. This leads to the conclusion that the outermost layers of the star — its atmosphere — are tenuous gas. The continuous background emission in the spectrum originates in deeper lying parts of the star, in the so-called photosphere.

Analysis of the absorption lines in the Sun's spectrum shows that the temperature of the gas in the solar atmosphere is close to 4700°K. In most other stars the atmospheric temperatures differ at most by a factor of 2—3, although there are extremely hot stars with temperatures approaching 100,000°K. Stars are classified in different spectral types or classes according to the predominant lines of the different elements in their spectra. The spectral type of a star is determined by the temperatures and pressures in its atmosphere. Table 1 lists the conventional symbols of the spectral types and the characteristic chemical elements of each type. The average temperatures measured from the spectral lines in each type are also given. Note that the Sun is a star of spectral type G.

TABLE 1

Spectral type	O	B	A	F	G
Characteristic lines (elements)	Helium ions	Neutral helium atoms	Hydrogen	Hydrogen and metal	Metal ions and atoms
Temperature, °K	30,000°	20,000°	10,000°	8000°	6000°

Spectral type	K	M	S	C
Characteristic lines (elements)	Metal atoms and various molecules	Titanium oxide molecules TiO	Zirconium oxide molecules ZrO	Cyanogen molecules CN
Temperature, °K	4000°		3000°	

Under the conditions prevailing in stellar atmospheres, all matter occurs in gaseous forms. This is atomic gas, and only the coldest stars (the so-called late-type stars) have atmospheres which contain molecules of two or more atoms. At temperatures above 4,000°, even the most stable molecules break up (dissociate). In addition to the chemical compounds listed in the table, the atmosphere of cold stars contains other molecules, but their effect on the spectra is not very pronounced.

The chemical composition of the stellar atmospheres is also found from their spectra. The relative abundance of the various elements in the stellar atmospheres is greatly different from that in the Earth's atmosphere. The most abundant element in virtually all the stars is hydrogen, whose content reaches about 60% of the total gas mass. The next abundant element is helium, whereas the heavier elements — oxygen, nitrogen, carbon, and metals — account for a mere 1—2% of the total gas mass. Very few stars show an increased content of the heavy elements and a relatively low abundance of hydrogen.

The physical conditions in stellar atmospheres are characterized by exceedingly low densities, as well as high temperatures. The number of gas particles in 1 cm^3 of the Sun's atmosphere is of the order of 10^{14}, which is hundreds of thousands of times less than the number of particles in the same volume near the Earth's surface. At temperatures of thousands and tens of thousands of degrees, gases are fairly opaque to optical radiation. A layer containing only a few grams of gas per 1 cm^2 of surface area is opaque even to continuous radiation. We simply cannot see into the star below this level: the light does not reach the observer from these interior parts. The mass of the gases accessible to direct visual observations is thus very low, of the order of a few grams per cm^2 of star's surface, which is negligible by terrestrial standards: there is more than 1 kg of air for each cm^2 of the Earth's surface.

There is no sharp boundary between the atmosphere and the deeper lying layers. As we penetrate deeper into the star, the gas density increases. Since the gas opacity increases very rapidly with the increase in temperature and density, the thickness of the atmosphere is a minor fraction of the star's radius. The surface of the star is roughly the level below which the observer cannot penetrate visually. The distance of this level from the center is regarded as the radius of the star. The stellar radii are inferred from the general laws of radiation, which will now be considered in some detail.

We know from daily experience that the hotter the body, the greater the amount of energy that it emits. This is applicable both to solids and to gases. Different bodies, however, differ in their emissivity. It has been established that the emissivity is related to the capacity of the object to absorb incident radiation. The best absorber of radiation is black color. An ideal object absorbing all the incident electromagnetic radiation is known as a blackbody. Real objects, however, always reflect a certain fraction of the incident energy, and they therefore only approximate to a certain extent to a blackbody. For example, a metallic sphere painted black is a fair approximation of the theoretical blackbody.

The energy E emitted per unit surface by a blackbody is a function of its temperature. This energy has been shown to depend on the fourth power of the absolute temperature T:

$$E = \sigma T^4. \qquad (6)$$

The proportionality coefficient σ in this relation, known as the Stefan—Boltzmann law, is equal to $5.7 \cdot 10^{-5}$ erg/deg$^4 \cdot$ sec $\cdot$ cm in CGSE units.

Blackbody radiation has another significant property. The dependence of the radiated energy on the wavelength (i.e., the spectrum of the radiation) is completely determined by the temperature. So as not to complicate the discussion, we will not write out the corresponding analytical expression of this dependence; the general behavior, however, is shown in graphical form in Figure 13, which plots the blackbody spectrum at various temperatures.

We see from the figure that as the temperature increases the wavelength λ_{max} corresponding to the peak emission (in unit wavelength interval) shifts toward shorter wavelength. The exact expression of the relation between λ_{max} and T, known as Wien's law, has the form

$$\lambda_{max} = \frac{0.28}{T}, \tag{7}$$

where λ_{max} is expressed in centimeters and T in °K.

The main contribution to the radiation from a body comes from the range of wavelengths around λ_{max}. This radiation determines the color of the emitting body. Thus, Wien's law in fact shows how the color of the body varies with its temperature. For example, at $T = 4500$°K we have $\lambda_{max} = 6.5 \cdot 10^{-5}$ cm, which correspond to red color. At temperatures near 6000°, we have $\lambda_{max} \approx$ $\approx 5.0 \cdot 10^{-5}$ cm, which is yellow light.

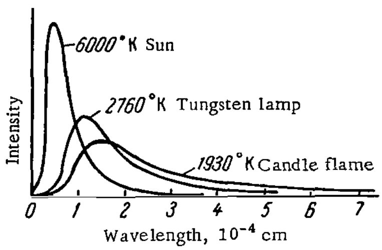

FIGURE 13. Energy distribution in the spectrum of a blackbody at various temperatures.

None of the stars is a perfect blackbody. However, since the photosphere readily absorbs radiation, its emission is regarded as closely approximating to the blackbody emission. This is a fairly crude approximation, since the observer sees at the same time radiation emitted in different layers of the star. As we go deeper into the star, the temperature increases. The temperature of the layers where the absorption lines form, i.e., in the star's atmosphere, is lower than in the photosphere, where the continuous emission originates. Moreover, the transparency of the photosphere is different at different wavelengths, and the radiation received at different wavelengths therefore corresponds to different temperatures. Nevertheless, the assumption that the stars are perfect blackbody emitters does not lead to substantial errors in the stellar radii.

The total quantity of energy radiated by a star in 1 sec is called stellar luminosity. It is generally designated by the letter L. If a star of radius R_* emits as a blackbody of temperature T, the luminosity is found by multiplying the energy E from relation (6) and the surface area of the star:

$$L = 4\pi R_*^2 \sigma T^4. \tag{8}$$

To find the radius from equation (8), we thus require the luminosity, as well as the temperature of the star. The luminosity is inferred from the quantity of energy reaching the observer, if the distance of the star is known.

For example, let us calculate the luminosity of the Sun; this will provide us with an order of magnitude estimate for many other typical stars. Direct measurements show that the energy reaching the Earth from the sun is $1.4 \cdot 10^6$ erg in 1 sec for each cm^2 of the surface. This figure is known as the solar constant. The same energy, however, will reach each second any surface element of 1 cm^2 area if it is located at the same distances from the Sun as the Earth and is perpendicular to the sun rays. Hence, the total energy radiated by the Sun in 1 sec is found if we multiply the solar constant by the surface area of the sphere whose radius is equal to the

Earth—Sun distance. The luminosity $L_\odot$ of the Sun (the subscript $\odot$ signifies that the quantity represents the Sun) is thus equal to $4\pi(1.5 \cdot 10^{-13})^2 \times \times\ 1.4 \cdot 10^6 = 3.9 \cdot 10^{33}$ erg/sec.

Using (8), we can now find the radius $R_\odot$. For $T_\odot$ we assume the approximate value of 6000°, which is obtained from Wien's law: we know that the Sun is a yellow star, and this color corresponds to temperatures of about 6000°. This is the temperature of the solar photosphere. A similar temperature is obtained for the photosphere of the Sun if it is determined from the slope of the curve which gives the energy emitted in the continuous spectrum as a function of wavelength. Using equation (8), we find $R_\odot \approx \approx 700{,}000$ km. Since the apparent radius of the Sun in angular units and the distance of the Sun are both known, $R_\odot$ can be calculated independently of the above assumptions. This alternative approach gives a very similar result for $R_\odot$.

For other stars equation (8) provides the only means for the calculation of the radius (except the few stars which form photometric binary systems; this phenomenon is described later on). The stellar temperature entering equation (8) is estimated from the continuous emission of the star; this is known as the effective temperature. The temperature determined from the absorption lines corresponds to the temperature of the atmosphere; it is about 15—20% lower than the effective temperature. The effective temperature of the Sun is 5800°, and the excitation temperature is 4700°.

When determining the luminosity of a star from the amount of energy reaching the observer, the star is compared with some source of known luminosity, e.g., the Sun. In these calculations, the energy flux is taken to be inversely proportional to the square of the distance from the source (see § 2). The distances to the nearest stars, distant not more than 100 light years from the Sun, are measured from their parallactic displacements. For other stars, however, the parallaxes are too small to be measurable, and their distances are not known; the luminosities therefore cannot be found either. In these cases more complex indirect methods of luminosity determination are used. They are based on the empirically discovered relation between the luminosity of a star of a given spectral type and certain characteristic features in its absorption lines. Since the spectra can be obtained for the farthest stars, the luminosity is determined from the spectrum, and then the distance is calculated from the luminosity.

The stars greatly vary in terms of luminosity. The Sun has a relatively low luminosity. Stars like the Sun and weaker stars, emitting 10^{33} erg/sec and less, are called dwarfs. The weakest dwarfs have luminosities which are thousands of times less than the luminosity of the Sun. On the other hand, there are giant and supergiants, which are hundreds and thousands of times more luminous than the Sun.

A certain relation is observed between stellar luminosities and temperatures. If the luminosity of each star (in units of $L_\odot$, say) and the spectral type (which is a characteristic of its temperature) are plotted in a graph, the resulting diagram has the form shown in Figure 14. For the first time, this diagram was constructed by Hertzsprung and Russell. The great majority of stars fall on the so-called main sequence in this diagram.

For these main-sequence stars, the luminosity, the radius, and the temperature are interrelated not only by equation (8). The temperature determines the luminosity, although not very accurately: to each temperature probably corresponds a certain fairly narrow range of permissible luminosities.

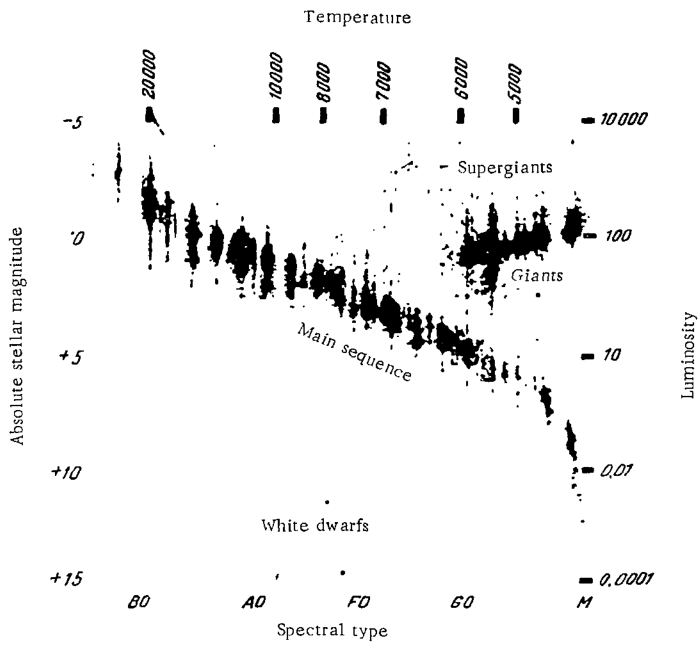

FIGURE 14. Spectrum—luminosity diagram.

Giants and supergiants fall in a distinct group in the top right-hand part of the diagram. They are relatively few compared to Sun-type stars, but because of their exceedingly high luminosity, they are visible over much greater distances than the dwarfs. We thus have an entirely false impression of the actual abundance of giants and supergiants.

The radii of stars fall between very wide limits. The dwarfs, which according to the diagram belong to spectral types G, K, and M, have radii of the order of $0.1\,R_{\odot} - 1\,R_{\odot}$. The giants and the supergiants of early spectral types (O, B, A) have radii of about $5\,R_{\odot} - 10\,R_{\odot}$. The temperatures of the K- and M-type giants are low and their high luminosity is associated with the extremely large surface area. Their radii are hundreds of times greater than the solar radius.

Approximately one out of every ten stars is a white dwarf, i.e., a star with a very small radius. Temperatures of white dwarfs are higher than the Sun's temperature (which explains their blue or white color), whereas their luminosities are one order of magnitude less. Using equation (8) we readily find that these stars are very small compared to the Sun, and are closer to planets in terms of their size.

In addition to stellar radius and luminosity, the mass is another highly important characteristic of stars. The masses of celestial bodies can be found only from observations of their attraction. We will start with a calculation of the Sun's mass. It is determined from the gravitational attraction experienced by the planets, and the Earth in particular. The gravitational attraction is mathematically expressed by Newton's law:

$$F = G\frac{M_1 M_2}{r^2}. \tag{9}$$

This is the equation for the attraction force F on a body of mass M_1 due to a body of mass M_2 at a distance r. The proportionality coefficient G is called the gravitation constant. If M_1, M_2, r, and F are expressed in CGSE units, we have $G = 6.67 \cdot 10^{-8}\,\mathrm{cm}^3 \cdot \mathrm{sec}^{-2} \cdot \mathrm{g}^{-1}$. The gravitation constant is determined in laboratory by measuring the attraction between bodies of known mass. Since the gravitational interaction is extremely small, these experiments are very complicated from the technical point of view.

If the relative velocity of two gravitating bodies is not too large, the bodies move around a common center of gravity. The square of the period of revolution P is proportional to the cube of the mean distance between the gravitating bodies:

$$P^2 = \frac{4\pi^2 a^3}{G(M_1 + M_2)}. \tag{10}$$

Taking $M_2 = M_\odot$, and using for M_1 some planetary mass, we obtain from (10) one of Kepler's laws of planetary motion, discovered from observations back in 17th century, before the advent of Newton's law of universal gravitation.* We will apply equation (10) to the Earth—Sun system, ignoring the effect of other planets. The period of orbital revolution of the Earth is 1 year, which contains $60 \cdot 60 \cdot 24 \cdot 365.25$ sec. Using this value for P and the mean distance of the Earth from the Sun $a =$ $= 150{,}000{,}000$ km, we find $M_\odot + M_{\text{Earth}} = 2 \cdot 10^{33}$ g.

Since the Earth's mass (calculated from the known gravitational acceleration g at the ground) is only $6 \cdot 10^{27}$ g, we find for $M_\odot$ a value of $2 \cdot 10^{33}$ g, which is 330,000 times the mass of the Earth.

We have thus obtained an order of magnitude estimate for stellar masses. The masses of other stars which are components of binary systems are found in the same way from equation (10). A binary system, or simply a binary, is a system comprising two stars, both moving around a common center of inertia under the force of mutual attraction. Binaries are a very common phenomenon in the star universe — more than 20% of all the stars are components of binary systems. The brightest of all the known stars — Sirius — is also a component of a binary. Many years of observations established that Sirius does not move in a straight path across the sky: its apparent trajectory is curved, probably due to the presence of a nearby star. This component is visible only

* The original expression of Kepler's law did not contain the planetary mass, but since it is invariably much less than the mass of the Sun the error was insignificant.

through a telescope. The two stars move in elliptical orbits around their common center of gravity with a period of about 50 years. The distance between the two stars is twenty times the Earth — Sun distance.

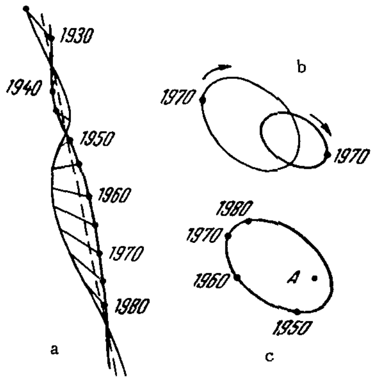

FIGURE 15. The apparent orbit of Sirius: a) The thick line traces the motion of the primary and the thin line is the trajectory of the minor companion. b) The motion of the two components around the common center of gravity (here a stationary center of gravity is assumed, whereas in figure a the motion of the center of gravity relative to the Sun is marked by the dashed line). c) The motion of the companion around the primary.

Binaries where the separation of the two components is large enough to make both stars observable are called visual binaries. Observations of visual binaries give directly the separation a and the period of revolution P, so that equation (10) can be applied to calculate the sum of the component masses. If the individual motion of each component is observed (and not only their relative motion), the mass of each star can be found.

If the separation of the two components is too small to resolve the system into two distinct stars through a telescope, the motion is studied spectroscopically. As each star moves around the center of gravity of the system, its velocity relative to the observer changes continuously. These changes produce a periodic shift of the spectral lines, associated with the Doppler effects. Stars which show periodic shift of spectral lines are known as spectroscopic binaries. The orbital periods of the components of spectroscopic binaries are fairly small, and only occasionally reach a few years. They generally do not exceed a few days, or even hours. Observations of spectroscopic binaries also provide valuable information on stellar masses, although this information is not as comprehensive as that for visual binaries.

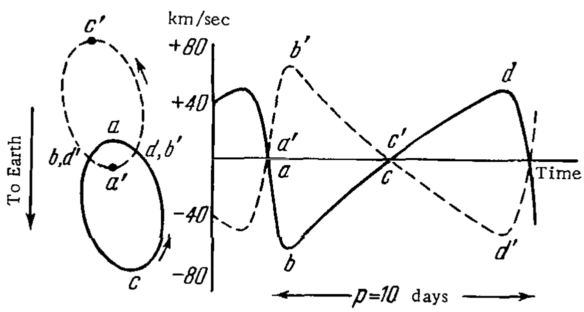

FIGURE 16. Motion in the spectroscopic binary ξ Ursae Majoris. The figure on the left shows the motion of the two components around the common center of gravity; right — the velocity curves of the components.

The geometry of the binary system can be such that the stars in the course of their orbital motion alternately hide each other from the observer.

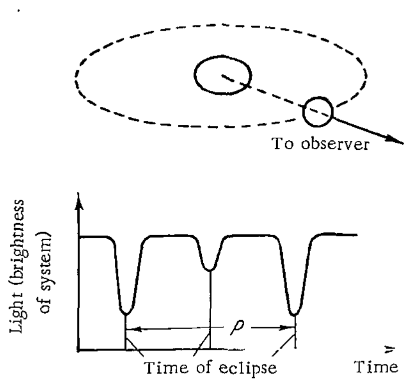

FIGURE 17. A schematic diagram of an eclipsing variable at the time of eclipse. The curve below shows the periodic variation of the apparent brightness of the star associated with the recurrent eclipses (because of the different luminosity of the two stars, the variation of brightness at the minimum depends on which of the two components is being eclipsed).

The observer's line of sight in this case lies in the orbital plane (the plane of motion of the two stars) or near it. These are the so-called photometric binaries or eclipsing variables: the recurrent eclipses lead to an observable fluctuation in the brightness of the system. Photometric binaries are not very numerous, since the orbital planes of the binaries are oriented at random relative to the terrestrial observer and only a tiny fraction of systems have their orbital plane set for an eclipse. Nevertheless, observations of eclipsing variables provide highly valuable information on stellar masses and also on structure of stars and their emission properties.

Thus our knowledge of the stellar masses is based entirely on observations of binaries. As a rule, however, the components of binary systems are similar in all respects to solitary stars. It therefore seems that the masses determined for binaries are fairly characteristic of all the stars. The masses of most stars are not much different from the solar mass. The highest mass is about $100 M_{\odot}$. The masses of red dwarfs are about 1/5—1/10 of the solar mass.

A definite relation between the luminosity L and the mass M of a star has been established: the luminosity is proportional to the cube of the mass, $L \sim M^3$. This is a statistical, and not an exact relation, and substantial deviations in either direction are possible. Moreover, only main-sequence stars follow this mass—luminosity relation. The mass—luminosity relation was also derived theoretically from modern models of stellar structure.

Binary systems are the simplest example of what is known as star systems. A star system is a group of stars with some common properties, which mainly pertain to the motion of the component stars and their position in space. Sometimes in a system comprising several stars (three, four, and more), the components are sufficiently close to one another to be linked by gravitational attraction, so that they move for an appreciable length of time in orbits around a common center of gravity. These systems are called multiple stars. A very interesting example of a multiple star is provided by Castor—the brightest star in the constellation of Gemini. Back in 18th century, Castor was identified as a visual binary, and in 19th century a third component was discovered, at a distance 12 times the mutual separation of the two pair components. Spectroscopic observations revealed that each of the three stars is a spectroscopic binary, so that all in all Castor is a sextuple system.

Another class of star systems — star clusters — contain much more numerous stars, but the bond between the components is much weaker than in multiple stars. Star clusters are divided into two groups: open clusters and globular clusters. An open cluster consists of several tens or hundreds of stars, all roughly at the same distance from the Sun; they all move approximately with the same speed in the same direction. One of the best known examples of this kind, the Pleiades cluster, contains over 250 stars, but only few of them are visible to the naked eye. Open clusters fill enormous volumes of space: their diameters are from 10 to 100 light years.

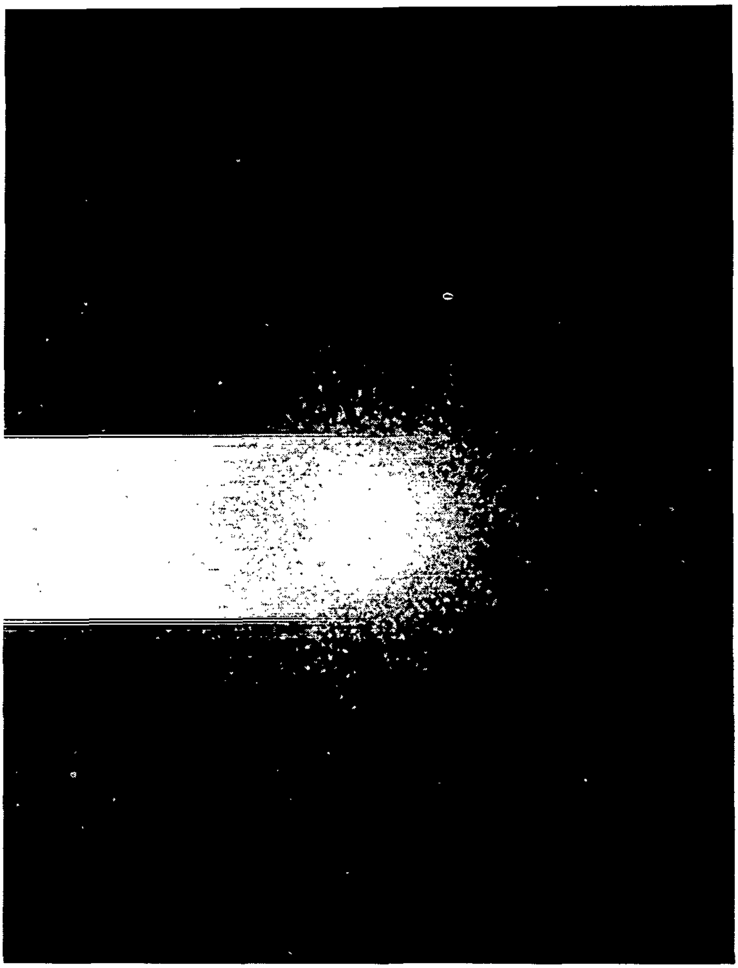

FIGURE 18. A globular cluster (M 13).

Globular clusters derive the name from their almost spherical shape. These are unusually compact groups of stars. A globular cluster contains hundreds of thousands of stars inside a volume with a diameter only a few times greater than the span of an open cluster. Therefore, near the center, the star images merge into one solid bright region. The number of stars per unit volume, say a cube 10 light years on a side, inside a globular cluster is tens of times higher than in the nearby space. The stars in globular clusters experience attraction due to the proximity of other stars in the cluster. This attraction is much stronger than in open clusters, where the interstellar distances are fairly large.

Star clusters of both kinds are part of a gigantic star system, known as the Galaxy. The Sun is only one of a multitude of stars in the Galaxy. The bulk of the constituent stars of the Galaxy is concentrated in the Milky Way. Telescopic observations of different parts of this luminous lane extending across the sky show that it consists of faint stars indistinguishable to a naked eye. The stars in the Galaxy (star counts show that their total number exceeds a hundred billion) lie mainly near a certain plane, which is called the galactic plane. The Galaxy is a highly flattened system of stars, and if it could be viewed edge on it would appear similar to the nebula photographed in Figure 19.

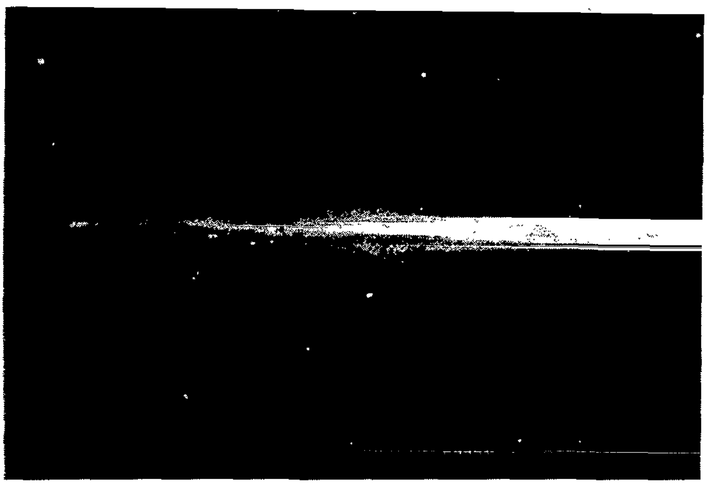

FIGURE 19. *A spiral galaxy viewed edge on (NGC 4565).*

When the Galaxy is viewed in plan (from above), we notice that the stars are not distributed uniformly over the disk, concentrating in the so-called spiral arms which stretch from the central nucleus. Although the Sun is located near the galactic plane and we cannot view the entire Galaxy in plan, these conclusions emerge from the results of star counts on photographs of different parts of the sky and especially from radio observations. The appearance of the Galaxy in plan is also inferred from the shape of other similar star systems. An excellent photograph of this kind is shown in Figure 20. The Sun lies in one of the spiral arms of the Galaxy, at a distance of 30,000 light years from the center. The radius of the Galaxy is about 50,000 light years.

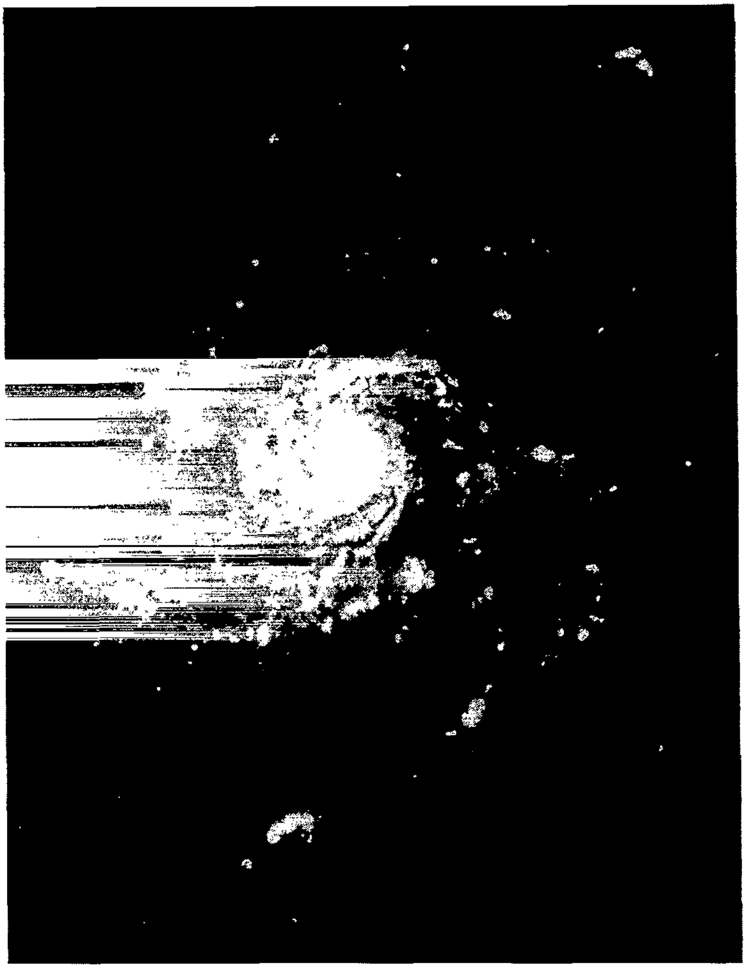

FIGURE 20. A spiral galaxy viewed in plan (NGC 5457).

The globular clusters also show a certain regular distribution in the galaxy. They are all concentrated in the central part, fairly symmetrically about the center, creating what is known as the spherical halo of the Galaxy.

All the stars in the Galaxy experience mutual gravitational attraction. Stars in the central part act as one whole on stars nearer the periphery. Because of this gravitational attraction, the peripheral stars, including the Sun, revolve around the galactic nucleus according to the same laws that drive the planets in their orbits around the Sun. In other words, their motion obeys Kepler's law, and the period of revolution is related to the distance from the center by equation (10).

The orbital period of the Sun around the center of the Galaxy is readily found if we know the distance of the Sun from the center and its velocity. Since objects lying outside the Galaxy do not take part in its rotation, the velocity of the Sun measured relative to the so-called extragalactic nebulae is equal to the velocity of its galactic rotation v. This velocity is found to be equal to 280 km/sec.* The Sun completes one circuit around the center of the Galaxy in a time $P = \frac{2\pi a}{v}$, where a is the distance from the center. Using the known values of a and v, we find P = 260,000,000 years. From equation (10) we can now readily calculate the mass M whose attraction is responsible for the orbital motion of the Sun in the Galaxy (the mass of the Sun is of course ignorable compared to M). The total mass of the stars lying inside the Sun's orbit in the Galaxy is thus found to be equal to 200,000,000,000 solar masses. The orbital period and hence the angular velocity vary with the distance from the galactic center, so that the Galaxy does not rotate as a rigid body.

Apart from stars, which are formations of dense matter, galaxies contain diffuse clouds of exceedingly low density. A cosmic explosion is generally accompanied by large-scale dispersion of matter, and a certain fraction of the diffuse interstellar matter has originated in a variety of catastrophic processes which converted part of the stellar mass into tenuous diffuse gas. A study of the state and motion of diffuse matter near the explosion focus provides highly valuable information about the explosion.

Stars greatly differ in their mean density, i.e., in the degree of concentration of matter. However, even in the low-density stars the material density is much higher than in the diffuse clouds. For comparison, let us consider some mean densities of stars, which are calculated as the ratio of the star's mass to its volume. They are easily found if the masses and the radii of stars are known. The mean density is the highest in white dwarfs. One cubic centimeter of a white dwarf contains about one ton of matter, and in some dwarfs the density is even one order of magnitude higher. This corresponds approximately to $10^{30}-10^{31}$ particles in each cm^3. The mean density of the gas in the Sun is 1.4 times the density of water, so that 1 cm^3 contains millions of times less particles than a comparable volume of a white dwarf. The largest stars, the red giants, have the lowest densities. Each cm^3 of a red giant contains a mere $10^{15}-10^{16}$ particles, which is thousands of times less than the number of particles in 1 cm^3 of air near the ground. But even this figure is tremendous compared to the density of diffuse matter.**

The interstellar medium contains only a few atoms per cm^3. The distribution of gas in the interstellar space is highly nonuniform. The

* In addition to their regular revolution around the galactic center, the stars also possess individual random motions. As a rule, these random velocities amount to 10—20 km/sec, which is much less than the velocity associated with the galactic rotation.

** These figures naturally do not characterize the state of matter in the cores of stars, where the density is much higher than the average. Thus, at the center of the Sun the density is over 100 g/cm^3.

bulk of the gas collects in enormous clouds with densities 10—15 times above the average. Clouds extend over hundreds and even thousands of light years and often have masses which are tens of thousands of times greater than the mass of the Sun. These clouds move with random velocities of a few kilometers per second. Besides gas, these clouds contain solid particles with diameters of 10^{-4}—10^{-5} cm. The concentration of these particles, which comprise the interstellar dust, is surprisingly low: there are no more than 40—50 dust particles in 1 km^3. However, because of its great extent, a dust cloud is either completely opaque to the radiation of stars behind it, or this radiation reaches the terrestrial observer drastically attenuated. The mass of the dust is no more than 1—2% of the total mass of interstellar matter.

The parts of the sky filled with relatively near gas-dust clouds appear as dark patches against the bright background peppered with thousands of stars. These patches are known as dark diffuse nebulae. The size, the density, and the distance of these diffuse clouds is inferred from the attenuation of the total light of the distant stars that they cause and from their absorption at different wavelengths. The interstellar gas absorbs radiation in definite spectral lines. Therefore the spectra of distant stars whose light passes through the interstellar gas on its way to the observer acquire additional absorption lines of interstellar origin. The intensity and the shift of these lines provide a clue as to the distribution of the absorbing matter in space and the velocities of the dust nebulae.

Radio observations supply us with a highly powerful tool for studying the interstellar medium. For the most part, the temperature of the interstellar gas is low, about 100°K, and it neither emits optical radiation nor absorbs it. Virtually all the hydrogen atoms at this temperature are in the ground state, and the energy of the optical quanta is simply insufficient for exciting the hydrogen to higher energy states. However, the ground state of the hydrogen atom consists of two very close sublevels, and the energy difference between these sublevels corresponds to the energy of a photon with a wavelength of 21 cm. Photons of this wavelength, which fall in the easily accessible part of the radio spectrum, are both absorbed and emitted by the interstellar hydrogen. Because of this lucky fact, the distribution of the cold diffuse gas clouds in the far parts of the Galaxy could be studied by radio observations, which readily penetrate into the far reaches of space which are normally obscured from optical observations by gas dust clouds: radio waves are hardly attenuated by dust.

Gas-dust clouds — the dark nebulae — lie in a fairly thin layer near the galactic plane, and together with stars they concentrate in the spiral arms. If we could view the Galaxy edge on, we would notice a dark lane across its middle, not unlike that visible in Figure 21. This is a photograph of a spiral star system viewed edge on. The dark lane in the plane of symmetry is created by a layer of interstellar matter which strongly absorbs the light of the stars in this plane.

There is no sharp boundary between the stars and the interstellar medium. The density of gas in the stellar atmosphere gradually decreases with the distance from the photosphere, eventually reaching the characteristic densities of the interstellar gas, although this occurs at tremendous distances. Thus, the outermost reaches of the Sun's atmosphere, at a distance approximately equal to the Sun's radius from the surface, still contain 10^4-10^5 atoms per cm^3, which is thousands of times denser than the interstellar medium.

FIGURE 21. A dark dust lane is clearly visible in this photograph of the spiral galaxy NGC 4594 viewed edge on.

The outer reaches of the Sun's atmosphere, the so-called solar corona, constitute in point of fact an extended gaseous envelope. Other stars are apparently also shrouded in gaseous envelopes of sorts. In what follows we will consider in some detail stellar envelopes produced by strong cosmic explosions. An explosion may generate a new envelope, or modify an existing one. The study of the gaseous envelopes of exploding stars provides highly valuable information on the nature of cosmic explosions. Because of their leading role, we will now consider in some detail the process of emission of the stellar envelopes.

One of the factors causing the emission of diffuse clouds is the excitation of the tenuous gas by radiation from a hot star. This process has been studied most comprehensively for the so-called planetary nebulae, which are in fact giant stellar envelopes.

Planetary nebulae generally appear as a bright disk or ring with a star at the center.* We can only see the projection of the nebula on the sky, and in fact this is a spherical object. The diameters of planetary nebulae are very large, of the order of one light year. The overall optical radiation from a planetary nebulae is considerably stronger than the radiation from the central star in the same spectral region. Hence, the radiation of the nebula cannot be regarded simply as scattered radiation of the nucleus star. The peculiar origin of the nebular emission is also confirmed by its spectrum. The most prominent lines in this spectrum are the emission lines of atomic hydrogen, atomic helium, and doubly ionized oxygen, whereas the spectrum of the central line shows no such lines. As we have already mentioned in the preceding, a spectrum of

* The first observers of these nebulae called them planetary because of a superficial similarity with the disks of some planets. They are absolutely unrelated to planets.

emission lines is characteristic of hot rarefied gas. How do the planetary nebulae emit, or in other words where does the radiated energy come from? This question was answered back in the 1930s. The theory of emission of planetary nebulae developed in that period played a highly important role in the treatment of emission from stellar envelopes and hence advanced our understanding of various effects associated with stellar explosions.

The central star in a planetary nebula has exceedingly high temperature, of the order of 100,000°. From Wien's law, most of the energy radiated at this temperature lies at wavelengths between $3 \cdot 10^{-6}$ and $5 \cdot 10^{-6}$ cm. The photons emitted in this region are more energetic by one order of magnitude than the photons in the optical region of the spectrum, and they are thus capable of ionizing hydrogen, helium, and other elements in the nebula. The atoms in the plasma produced by the stellar radiation undergo recombination.

An ion combining with an electron sometimes produces an atom in the ground state and sometimes gives an excited atom. An excited atom dropping to the ground state emits photons in certain spectral lines. The source of the emission energy in planetary nebulae is thus the radiation of the hot central star in the invisible ultraviolet region of the spectrum. This short-wave radiation is degraded by the nebula into optical radiation, mostly at the frequencies of spectral lines. This process is known as fluorescence.

Fluorescence occurs freely only if an atom during the lifetime of its excited state does not collide with another particle (or photon), which may excite it to a still higher energy state or remove some of its energy. The nebular spectrum shows that atoms do indeed drop to lower energy states. This implies that the concentration of atoms in the nebula and the number of photons per unit nebula volume are sufficiently small. These conditions are also satisfied in stellar envelopes and in interstellar gas. Therefore, the gas around the hot stars (types O and B) is expected to emit a line spectrum, not unlike that of planetary nebulae. If, however, the temperature of the star is insufficiently high, the stellar radiation will not excite emission from the gas around the star.

The luminous patches in the interstellar medium are called bright diffuse nebulae. Apart from fluorescing nebulae, there are also other forms of bright nebulae. They are visible in the light of nearby stars which is scattered and reflected by dust particles. These nebulae are also identified from their spectrum, which in all respects is analogous to the spectrum of the illuminating star.

When we see an emission spectrum we naturally conclude that the luminous gas emits through excitation and ionization, followed by recombination of atoms. One of the reasons for ionization is completely unrelated to the radiation of the hot star. Consider, for example, two colliding gas clouds which approached each other with a velocity of several tens of kilometers per second. Part of the kinetic energy of the clouds is expended in ionization of atoms by collision. Recombining, these atoms emit energy in the line spectrum. In this case, the emission energy is derived from the kinetic energy of the gas.

Investigation of emission lines in the spectra of nebulae and stellar envelopes yields valuable information on the emission source (in particular, the temperature of the exciting star, if the emission is associated with fluorescence) and on the state of the emitting gas. For example, the gas density is inferred from the brightness of the nebula, as determined by the number of photons emitted in various spectral lines. These photons originate in recombinations, i.e., in encounters between ions and electrons. The frequency of these encounters naturally depends on the velocity of the particles (the gas temperature) and their concentration in the gas. The number of photons emitted per unit volume of the nebula per second is determined from the line intensity. This function is also calculated theoretically from known atomic parameters for various gas temperatures and densities. Comparison of the observed value with the theoretical results fixes the temperature and the density of the nebular gas. This method gave densities of 10^3-10^4 atoms per cm^3 for planetary nebulae, whereas the bright diffuse nebulae were found to contain $1/20-1/100$ of this number. The temperature in either case was found to be around 10,000°K.

We considered various forms of matter in the Galaxy. The bulk of matter — some 90% of the total mass — is collected in stars, whereas the balance is distributed in dark and bright nebulae.* Outside the limits of our Galaxy, matter occurs in the same fundamental forms. The nearby space contains other star systems — galaxies — some of which are highly reminiscent of our system, whereas others are markedly different. We do not know of any galaxy in the Universe where the stellar and the diffuse forms of matter are not predominant.

The average distance between galaxies is of the order of several million of years. The part of the Universe accessible to observations with the largest modern telescopes — this part is called the Metagalaxy — contains over ten billion galaxies. The galaxies show a definite tendency to cluster. Groups or clusters of galaxies are observed, containing hundreds or thousands of members. The intergalactic space is not perfect vacuum, but the intergalactic matter is extremely tenuous. The astronomers' opinions vary as to the density of the intergalactic medium, but it can hardly exceed some ten atoms per cubic meter. A simple calculation, that the reader himself can perform, shows that the bulk of matter in the Metagalaxy is concentrated in the individual galaxies, so that within the Metagalaxy the stellar form of matter predominates. However, recent discoveries of new phenomena in the Metagalaxy, such as quasars and explosions in galactic nuclei, seem to suggest that the Universe contains other, hitherto unknown, forms of matter.

§ 4. THE ENERGY OF COSMIC BODIES

Let us now consider the different forms of energy in the Universe. The importance of this topic for cosmic explosions is self-evident, since all the processes in the Universe, explosions included, are associated with conversion of energy from one form to another. Without knowledge of the exact form of energy in celestial bodies, we cannot understand why and how they explode.

* A certain minor fraction of matter is apparently concentrated in dark large bodies, such as the planets of our solar system.

Energy so far has not been defined in terms of any other simpler concepts in science. Anyone who has studied some physics will be familiar with the different forms of energy and the techniques for calculating the quantity of energy. Despite the great variety of forms, however, energy almost always can be simply classified either as potential or as kinetic, although this is not immediately obvious for some forms of energy (e.g., the energy of the magnetic field).*

Energy which depends only on the relative position of the objects is called potential, irrespective of whether we are concerned with the position of a planet relative to the Sun, the position of a gas particle among other particles, or the position of an electron in the atom. Kinetic energy, on the other hand, is associated with motion of bodies or particles. When the motion of a system of objects or particles is observable on fairly large, macroscopic scale, we have kinetic energy in the usual sense of mechanics. The motion of photons, on the other hand, also involves kinetic energy. Microscopic motions, such as the motion of the molecules in a solid, the motion of particles in an atom, and the motion of atoms in a gas, again involve certain kinetic energy. The kinetic energy of microscopic motions is generally regarded as the inner energy of the body. Another component of the inner energy is produced by particle interactions, and as such it is a potential energy.

It has been firmly established by physical experiment that energy can neither disappear nor be created from nothing. This is the essence of the principle of conservation of energy. If we consider a physical system, ignoring all possible interactions with other systems, then the sum of the potential and kinetic energy in this isolated system remains constant. However, no system can be completely isolated from interaction with other objects, so that the above statement is true only to a degree. Any physical system exchanges energy with other objects; this transfer of energy is sometimes ignorable, but then again sometimes it is not. For example, when we study the motion of planets around the Sun, we can rightly ignore the attraction of the fixed stars, as they are very distant. The motion of the solar system on the whole, however, is definitely influenced by the constituent stars of the Galaxy: as we know, the resultant attraction of these stars makes the Sun move in an orbit around the center of the Galaxy. Other forms of interaction are also known: a certain fraction of stellar radiation reaches the planets, they are exposed to the impact of cosmic rays, etc. The principle of energy conservation thus does not maintain that the energy remains constant in any system, but rather that the energy cannot disappear or be created.

As we know, the bulk of matter in the accessible part of the Universe occurs in the form of stars. We will therefore start our discussion of the form of energy in the Universe with an estimate of the energy contained in stars. Ignoring the outermost cold layers, we know that stars are made up of completely ionized gas of ions and electrons. Since hydrogen is the prevailing chemical element, the number of particles in the stellar gas is roughly double the number of hydrogen atoms: remember that each atom is ionized into two particles, a proton and electron. The molecular

* A detailed and fairly popular discussion of the forms of energy will be found in the Feynman Lectures on Physics, Addison—Wesley Publishing House. 1963.

weight of hydrogen is 1, and the average molecular weight of the stellar matter is therefore close to 1/2.

The gas particles move and their total kinetic energy creates a thermal energy which is confined inside the star. Each particle also experiences attraction due to other particles, according to the law of gravitation. This attraction is much more significant than the electrostatic interaction. Although the plasma is made up of charged particles, on the whole it is neutral: in any sufficiently large volume the number of positively charged particles is equal to the number of negatively charged particles. Hence, any volume of the plasma does not experience electrostatic forces. This does not apply to gravitation, however: the attraction experienced by any particle is the sum total of the attraction forces due to all the other particles. The potential energy of a star is thus almost entirely due to gravitation.

Let us assess the potential energy of a star. It is determined by the work that is required in order to "break up" the star into constituent particles and move them apart to such distances that the gravitational attraction is negligible. In this state, the potential energy of the system can be regarded as zero. Since work must be done in order to bring the system to a state of zero potential energy, the potential energy of a body is defined as a negative quantity. The energy doing the work, e.g., thermal energy, is considered positive.

The work against the attraction forces required to "break up" a star is calculated by standard methods of calculus. The principle of these calculations is fairly simple. We successively detach thin skins of matter from the star and calculate the work done to remove each of these skins to infinity; after that the work for all the skins is added up. Exact calculations are possible only if the structure of the star is known. Since the distribution of stellar density is not known, we can only obtain an order of magnitude estimate of the potential energy.

Suppose that a layer containing half the total stellar mass is detached from the star and moved to infinity. We further assume that the mass of this detached layer is formed into a sphere similar in all respects to the remaining mass, i.e., we regard the star as made up of two equal spheres with their centers at a distance $\frac{R_*}{2}$, where R_* is the radius of the star. The gravitational interaction between these two components is described by equation (9), where we should take $M_1 = M_2 = \frac{M_*}{2}$ (M_* is the stellar mass) and r $\frac{R_*}{2}$. There is, however, still another difficulty: we can calculate work of a constant force, whereas here the force varies with the distance between the bodies.

Since the force rapidly diminishes with the distance, we will only calculate the work needed to double the distance between the mass centers— from $\frac{R_*}{2}$ to R_*. The force will be assumed to remain constant in the process, and in this way we compensate to a certain extent for the work neglected at distances greater than R_*. The potential energy U is then found in the form

$$U = -F\frac{R_*}{2} = -G\frac{\left(\frac{M_*}{2}\right)^2}{\left(\frac{R_*}{2}\right)^2}\frac{R_*}{2} = -G\frac{M_*^2}{2R_*}. \tag{11}$$

Let us use equation (11) to calculate the potential energy U of the Sun. We have previously obtained $M_\odot = 2 \cdot 10^{33}$ g and $R_\odot = 7 \cdot 10^{10}$ cm. The potential energy of the Sun is thus found to be of the order of 10^{48} erg. An exact calculation gives a higher potential energy: after all the "fragmentation" of a star does not terminate with its breakup into two equal spheres.

Note that U is not to be regarded as the energy reserve of a star: conversely, this is the energy that is released when the star is formed by compressing a diffuse cloud of matter into a compact star. As the radius of a star is halved, the potential energy U is doubled in absolute value, but remains negative, $2U$. An energy of U has thus changed to other forms of energy. Theoretically, the reserve of potential energy of a star is unlimited, since its radius can be made arbitrarily small (again theoretically). In fact, however, a star will only shrink so far, and the energy released in the process remains finite.

A highly important proposition, known as the virial theorem, is also proved by standard calculus methods. According to the virial theorem, the potential energy U is related in a definite way to the thermal energy T in any stable system of particles which are isolated from all external forces and interact according to the inverse square law of equation (9):

$$2T = -U, \tag{12}$$

or in other words double the thermal energy is equal to minus the potential energy. A star can be treated with fair accuracy as a system of particles isolated from a-1 external interactions, since other stars hardly affect the state of the individual particles in a star. Equation (12) is therefore applicable to any star, and it is applied to estimate the thermal energy of stars. For the Sun the thermal energy, like the potential energy, is of the order of 10^{48} erg.

The thermal energy of a star is distributed between all the constituent particles, but not in equal measure. The temperature in the stellar interior is higher than at the surface, and the kinetic energy of particles is therefore higher in the core. Nevertheless, it is interesting to calculate the average energy per particle. This average value will enable us to find the average temperature of a star.

To calculate the number of particles in the Sun, we assume that it is made up of pure hydrogen. The number of atoms in the Sun is thus obtained as the ratio of its mass to the mass m_H of a hydrogen atom* ($m_H = 1.65 \cdot 10^{-24}$ g). Since $M_\odot = 2 \cdot 10^{33}$ g, we find $N \approx 10^{57}$ particles. Since the hydrogen in the Sun is completely ionized, we obtain for the total number of particles a figure of about $2 \cdot 10^{57}$.

The ratio of the total thermal energy of the Sun to the number of constituent particles is of the order of 10^{-9} erg. The average energy of a

* The mass m_H is found by dividing the mass of one gram-atom by the Avogadro number.

gas particle is related to the temperature of the gas by equation (1), which for $\frac{m\bar{v}^2}{2} \approx 10^{-9}$ erg gives an average temperature of about $5 \cdot 10^6$ °K for the solar interior. At the center the temperature is higher than the average, and detailed calculations show it to be near $15 \cdot 10^6$ °K.

The gas pressure in the Sun's interior can also be estimated without difficulty. To this end we use equation (2), where the average value of n — the number of particles in 1 cm^3 of solar matter — is calculated as the ratio of the total number of particles N to the volume of the Sun, $4\pi R_\odot^3/3$. The calculations give for n a value of around $4 \cdot 10^{23}$ particles /cm^3, and for the pressure we find $P \approx 3 \cdot 10^{14}$ dyne/cm^2, i.e., hundreds of millions of atmospheres.

This enormous pressure (according to terrestrial concepts) in the stellar interior may seem quite puzzling and incredible. It is evident, however, that lower interior pressures simply would not support the enormous bulk of a star, and the entire mass would simply "collapse" to the center under its own gravitation. Indeed, let us consider the analogous situation on the Earth. We know that weight is associated with the attraction of objects toward the center of the Earth. A pressure of one atmosphere is the weight of air per unit surface area; it amounts to about 1 kg per 1 cm^2. As we go deeper into the Earth, the weight of the overlying rock and soil strata adds up to the weight of the atmosphere. The pressure rapidly increases as we move from the surface into the Earth's interior, and in the central regions it reaches a few million atmospheres.

A similar case obtains in a star. At any level, the pressure is determined by the weight of the overlying gas layer. Since the mass of a star is enormously large, the weight per unit surface area of a layer halfway between the center and the surface is much greater than the corresponding figure on the Earth. The stellar gas can support this weight only by remaining at exceedingly high pressure, which is determined both by high density and high temperature.

Stars of different types have different interior temperature and pressure. The values calculated for the Sun are typical of the average stars. In the coldest stars, the interior temperature is somewhat lower, but it is nevertheless several million degrees. A gas at such extremely high temperatures should emit tremendous quantities of energy. The emitted photons are absorbed by the atoms and are then again re-emitted, gradually diffusing into the outermost strata of the star. Having reached the surface, the photons escape into space and a certain fraction of this flux is detected as star light.

Apart from normal material gas, made up of conventional particles, the star thus holds a multitude of photons, which constitute a so-called "photon gas." Let us calculate the total energy of the photon gas, i.e., the energy of the radiation in the stellar interior.

The emissive properties of the stellar interior are very like those of a blackbody. Any volume element in the stellar interior absorbs all the incident energy, and the energy emitted from the surface enclosing the volume element should therefore be equal to the incident energy. Any surface in the interior of a star thus emits as the surface of a blackbody.

Consider a spherical surface of infinitesimal radius r, enclosing a certain volume in the stellar interior. Inside this volume, and hence on the hypothetical spherical surface, the temperature T is virtually constant. The quantity of energy emitted from the spherical surface in 1 sec is equal to $4\pi r^2\sigma T^4$. The photons remain inside the sphere only as long as it takes them to traverse the diameter with the velocity c.* A photon moving in the radial direction will traverse the sphere in a time $\frac{2r}{c}$. But photons are emitted by the surface in all directions and move along the chords, as well as along the diameter. More detailed calculations show that an average distance traversed by a photon inside a volume element is $\frac{4}{3}r$. Hence, at any time, the radiant energy locked up in the sphere is equal to $4\pi r^2 \cdot \sigma T^4 \cdot \frac{4}{3}\frac{r}{c}$.

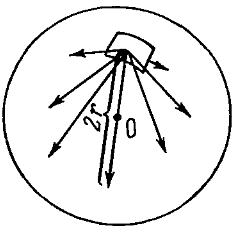

FIGURE 22. Diffusion of radiation inside a spherical volume.

Dividing this figure by the spherical volume $\frac{4}{3}\pi r^3$, we find the radiation density, i.e, the quantity of radiant energy in 1 cm^3. We see that it is equal to $\frac{4\sigma}{c}T^4$. The total quantity of radiant energy E_{rad} in stellar interior can be calculated only from known values of T for each volume element. A correct order of magnitude of E_{rad} is obtained, however, when we multiply the radiation density corresponding to the average temperature of the star by its volume:

$$E_{rad} \approx \frac{4\sigma}{c} T_{av}^4 \cdot \frac{4}{3}\pi R_*^3. \qquad (13)$$

In (13) R_* is the star radius. Inserting for T_{av} and R_* the temperature and the radius of the Sun, we find $E_{rad\odot} \approx 10^{43}$ erg. The radiant energy enclosed in the solar interior is thus much less than its thermal energy. This is also true for the other stars.

A star continuously emits energy from its surface, and this emission is essentially the radiation of hot gas sustained by the thermal energy of the star. The thermal energy contained in a star should thus be continuously replenished, since otherwise the entire reserve will soon be spent as radiation. For example, the Sun radiates at a rate of $3.9 \cdot 10^{33}$ erg per second, so that its thermal energy will not last more than thirty million years. And yet we know from geological findings that the level of solar radiation remained approximately constant for over a billion years.

The energy lost through radiation clearly cannot be replenished to any noticeable degree at the expense of the potential energy. Even if the Sun originated from a highly tenuous cloud, whose potential energy was close to zero, the energy released due to the compression of this cloud to a radius equal to the present-day size of the Sun would be 10^{48} erg, which again is not enough to sustain the solar radiation for long.

A similar situation is observed for other stars, at least for those which are like the Sun in terms of mass and luminosity. Stars therefore must contain other sources of thermal energy, which hitherto remained latent. We will be able to assess these energy reserves after acquainting

* If some of the photons are absorbed, they are replaced by the same number of photons emitted inside the volume element. The absorption is therefore ignored in this discussion.

ourselves with one of the basic concepts of the theory of relativity, namely the equivalence of mass and energy.

With the advent of the relativity theory, we had to reject the traditional notion of the conservation of mass. Mass was found to depend on the velocity of the object. The variation of the mass as a measure of inertia follows directly from the conservation of linear momentum. Momentum is created only when an external force is applied. If, in particular, a constant force F acts on a body during a time t, it will change the momentum mv (m is the mass of the body and v its velocity) from the initial value of $(mv)_0$ to some value $(mv)_t$, such that

$$(mv)_t - (mv)_0 = Ft. \tag{14}$$

One of the basic postulates of relativity is that no motion can exceed the velocity of light c. If the initial velocity of the body was close to c, no amount of force will markedly increase its velocity. The momentum, on the other hand, will steadily increase, and if the force is applied for a sufficiently long time $(mv)_t$ can be made as large as desired. Hence, for velocities close to c, the mass of the body increases, i.e., the inertial resistance of the body to further acceleration grows. The faster the body moves, the higher is its inertia. The dependence of the mass on the velocity of motion is described by the relation

$$m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{15}$$

which is derived in standard textbooks on relativity.

Here m_0 is known as the rest mass: this is the mass of the body when its velocity is zero. The dependence of mass on velocity is noticeable only for v of the order of tens of thousands of kilometers per second. Even at the normally tremendous velocities of v = 1000 km/sec, when $\frac{v^2}{c^2} = \frac{1}{90000}$, m differs from m_0 by a mere 0.002%. Therefore the change in mass due to motion is felt only in very fast particles, which move with velocities close to the velocity of light, e.g., particles accelerated by the magneitc field in a synchrophasotron or in cosmic ray particles. If $\frac{v}{c}$ is much less than unity, the relation (15) between mass and velocity is approximately written in the form

$$m = m_0\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = m_0 + \frac{m_0 v^2}{2c^2}. \tag{16}$$

For $\frac{v}{c} \leqslant \frac{1}{10}$ this relation is accurate to within 0.005%.* We thus see that the mass of a moving body increases by an amount equal to its kinetic

* Since for sufficiently small $\frac{v}{c}$, $\frac{v^4}{c^4}$ is very small compared to $\frac{v^2}{c^2}$, we have $\left(1-\frac{1}{2}\frac{v^2}{c^2}\right)^2 = 1 - 2\cdot\frac{1}{2}\frac{v^2}{c^2} + \frac{1}{4}\frac{v^4}{c^4} \approx 1 - \frac{v^2}{c^2}$. Therefore $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \approx \frac{1}{\sqrt{\left(1-\frac{1}{2}\frac{v^2}{c^2}\right)^2}} = \frac{1}{1-\frac{1}{2}\frac{v^2}{c^2}} = 1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{1}{4}\frac{v^4}{c^4} + \ldots \approx 1 + \frac{1}{2}\frac{v^2}{c^2}$. Here we have used the standard formula for the sum of a geometric progression (with a common ratio of $\frac{1}{2}\frac{v^2}{c^2}$).

energy divided by c^2. If we write Δm for this increase in mass, $m - m_0$, we obtain a highly important relation

$$\Delta m = \frac{E}{c^2}. \tag{17}$$

We have restricted the discussion to the case of velocities v much less than c only to simplify the mathematics. The above relation is true for any $\frac{v}{c}$.

Equation (17) establishes a relation between mass and energy. Although it was derived for the case when E is a kinetic energy, the result is in fact applicable to energies of other forms as well: in short, a certain mass always corresponds to any given amount of energy. Energy and mass are equivalent, i.e., a change in energy is always proportional to a change in mass.

Because of the equivalence of mass and energy, a body has a certain characteristic energy even if it does not move. This rest mass, m_0, is equivalent to a certain rest energy $E_0 = m_0c^2$. Under certain conditions the rest energy E_0 can be converted to other forms of energy, e.g., radiation. This conversion is known as annihilation ("destruction"). Annihilation processes can be actually observed, and this constitutes a brilliant confirmation of the general conclusions of relativity concerning the equivalence of mass and energy. A common event of this kind is the annihilation of an electron when it encounters another particle, known as a positron. These particles have the same mass and electric charge of the same magnitude, but they are oppositely charged. Combining, an electron and a positron annihilate, cease to exist, and what remains is radiation. The energy of the photons produced by this annihilation is $1.6 \cdot 10^{-6}$ erg, which is precisely equal to double the electron mass $2.9 \cdot 10^{-28}$ g multiplied by the velocity of light sqnared.

The mass of a star M_* corresponds to a rest energy M_*c^2. For the Sun we have $M_\odot c^2 = 1.8 \cdot 10^{54}$ erg. This figure is hundreds of thousands of times greater than the thermal energy of the Sun. This is quite understandable, since the average velocities of the constituent particles in the Sun are small compared to the velocity of light. The product M_*c^2 determines the total reserve of stellar energy. However, only some of this energy can be converted to other forms in stars.

Total annihilation is possible only if particles and corresponding antiparticles are present in equal amounts. The evidence on hand seems to indicate that the Sun and other stars in the Galaxy contain only negligible quantities of antimatter, if any. Therefore, the release of energy by thermonuclear fusion energy which is considered in what follows involves a minute decrease in stellar mass.

Let us find the change in the mass of a star following the release of energy in the so-called proton-proton cycle. According to modern concepts, the reactions of this cycle constitute the main source of energy in the Sun and other similar stars. This cycle produces helium neclei from hydrogen nuclei — protons.

The nuclei of all atoms, except hydrogen, are composite systems, made up of positively charged protons and electrically neutral particles of almost the same mass, neutrons. We recall that the helium nucleus (also known as an alpha particle) consists of two protons and two neutrons. The particles in a nucleus are bound by special nuclear forces, which are effective over very short ranges only. To remove a proton from an alpha particle, work should be done against the nuclear forces, i.e., some energy must be expended. This energy is known as the binding energy of a proton in an alpha particle. When two protons and two neutrons combine into an alpha particle, an energy equal to their binding energy is released.

Without going into the conditions necessary to ensure formation of alpha particles from protons, let us consider the different stages of this cycle. If in an encounter between two protons (we will write p in short for a proton) the particles manage to approach each other to a sufficiently small distance, a deuterium nucleus (heavy hydrogen D^2) is formed, which is made up of a proton and a neutron. The neutron in the D^2 nucleus is obtained from the proton. The conversion of a proton into a neutron is associated with a release of a positively charged particle, a positron e^+ (we have mentioned it before), and another particle of zero electric charge, the neutrino ν. The combination of two particles moreover involves a release of a photon of very high energy, a γ quantum. The reaction of formation of D^2 is thus schematically written in the form

$$p + p \rightarrow D^2 + e^+ + \nu + \gamma. \tag{I}$$

The positron e^+ is subsequently annihilated in an encounter with a free electron, and two γ quanta are created.

The deuterium nucleus meeting a proton will form a new particle, a light helium isotope He^3 containing two protons and a neutron. This reaction is also accompanied by a release of a γ quantum:

$$D^2 + p \rightarrow He^3 + \gamma. \tag{II}$$

The product nucleus He^3 may take part in various reactions, but the most probable of these in stars is the reaction between two He^3 nuclei. It proceeds according to the following equation:

$$He^3 + He^3 \rightarrow He^4 + p + p + \gamma, \tag{III}$$

i.e., it produces an α particle (He^4), two protons, and a γ quantum.

The three reactions of the proton-proton cycle thus convert four protons into an α particle, several γ quanta, and a neutrino. The rest mass of an α particle m_{He^4} is less than the mass of four protons $4m_p$. The difference $4m_p - m_{He^4}$ gives the mass which is equivalent to the energy released in the process, and the ratio $\frac{4m_p - m_{He^4}}{4m_p}$ is the fraction of the mass converted to energy by the proton-proton cycle. It is equal to $\frac{4 \cdot 1.0081 - 4.0039}{4 \cdot 1.0081} = 0.007$ (the masses are expressed in atomic mass units). We see that a complete proton-proton cycle reduces the mass only by seven tenths of a percent of the original hydrogen mass.

Thermonuclear reactions occur in a plasma only under certain conditions, and these do not obtain either in the outer layers of a star or in the interstellar medium. Not just any encounter between two protons will produce a deuterium nucleus. Protons are electrically charged particles of like charge, and they thus repel one another; the electrostatic repulsion increases as the particles draw nearer. Each proton is thus surrounded by a barrier which prevents it from freely combining with another positively charged particle. We say that a proton exists in a potential barrier. On the other hand, the nuclear attraction forces fall off rapidly with distance, and it is only at interparticle distances of the order of 10^{-13} cm and less that they exceed the electrostatic repulsion forces. A deuterium nucleus is thus formed only if the relative velocity of the colliding protons is large enough to overcome the potential barrier. One of the protons penetrates the electrostatic barrier and is converted into a neutron inside the effective range of nuclear forces. Particles of like charge have to overcome a similar potential barrier in all other thermonuclear fusion reactions.

If the plasma temperature is not very high, 3—5 million degrees say, there are no protons with kinetic energies sufficient for penetrating the barrier and the proton-proton reaction does not take place. At temperatures near 15 million degrees, a fairly large proportion of protons are sufficiently energetic for overcoming the electrostatic potential barrier and producing a composite nucleus. Thus, one of the necessary conditions of a proton-proton cycle is that the plasma be extremely hot, with temperatures of the order of $15 \cdot 10^{6}$ °K. On the other hand, to ensure sufficiently frequent encounters between protons, the concentration of these particles should be fairly high. Both these conditions are naturally met in the interior of the Sun and other similar stars. Hydrogen is the prevalent element in the Sun, and it is completely ionized in the solar interior. The number density of protons in the solar interior is about 10^{26} cm^{-3}, and the temperature reaches 15 million degrees. The proton-proton reactions which take place in the solar interior sustain the observed energy output of $3.9 \cdot 10^{33}$ erg/sec.

The energy in the proton-proton cycle is released mainly in the form of γ quanta, which interact with the stellar matter and are thus converted to heat. Part of the energy is removed with the neutrinos. The neutrino has an extraordinary property: it hardly reacts with other particles, and it thus penetrates without meeting any obstacle through the entire bulk of the Sun. The flux of neutrinos from the Sun and from other stars propagates unhindered through the interstellar space and all the material objects in it, the Earth included, without producing any noticeable effect.*

The conversion of the entire hydrogen reserve of the Sun into helium will produce $2 \cdot 10^{33} = 0.6 \cdot c^2 \cdot 0.007 \approx 7.6 \cdot 10^{51}$ erg. Given the current rate of energy radiation, this reserve will sustain the Sun for $2 \cdot 10^{18}$ sec, i.e., tens of billions of years.

All the above forms of energy are in fact the internal energy of a star. Stars, however, also possess a certain mechanical energy which is associated first with their motion as one whole and second with their interactions with other bodies in the stellar system.

Observations of stellar spectra show that numerous stars spin around their axis. The equatorial velocities of stars due to their axial rotation are equal to tens and sometimes even hundreds of kilometers per second. Slow rotation leading to equatorial velocities of a few km/sec is difficult to detect from spectroscopic observations. It is a known fact, however,

* Recent experiments detected neutrinos of solar origin (in very small numbers).

that the Sun spins with a linear velocity of 2 km/sec at the equator. It would thus seem that even stars whose spectra do not show any detectable rotation also spin with low velocities of this order of magnitude.

The rotation velocity is different at different points in the stellar interior: it depends on the distance of the particular point from the spin axis. Therefore, to obtain an accurate estimate of the kinetic energy of a star, we require the distribution of matter inside the star. If we assume that the concentration of stellar matter toward the center is not very pronounced, an order-of-magnitude estimate of the kinetic energy of rotation can be obtained by taking the mass of the whole star to be concentrated at some fixed distance from the spin axis, $\frac{R_*}{2}$, say, where R_* is the radius of the star. If the angular rotation velocity is ω, the linear velocity of the mass in this model is $2\pi\frac{R_*}{2}\omega$, and for the kinetic energy of rotation $E_{\rm rot}$ we find

$$E_{\rm rot}=\frac{1}{2}M_*(\pi R_*\omega)^2. \qquad (18)$$

Since the linear velocity at the Sun's equator obtained from direct observations is $2\pi R_\odot\omega_\odot = 2\cdot 10^{-5}$ cm/sec, its rotation energy is $E_{\rm rot\odot}\approx 10^{43}$ erg. This is very small compared to the thermal energy. For a fast spinning star, with equatorial velocities of 300 km/sec, say, the energy of rotation may reach 10^{47} erg. This figure approach the theoretically permissible maximum, since at much higher velocities the star will be torn apart by the centrifugal forces. Hence, even in the fastest spinning stars the rotation energy is smaller than the thermal energy.*

The mechanical energy of stars in binary systems may be very substantial. Consider, for example, two stars of equal mass ($M_1 = M_2 = M_\odot$), which revolve around a common center of gravity; the distance a between the stars is equal to the Earth—Sun distance. Since the velocity V of both stars is the same, their kinetic energy is equal to $\frac{1}{2}(M_1+M_2)V^2 = M_\odot V^2$. The velocity V is related to the orbital period by the relation $P=\frac{2\pi\cdot\frac{a}{2}}{V}=\frac{\pi a}{V}$. Using this relation and equation (10), we find $V^2=\frac{GM}{2a}$, and the kinetic energy is $\frac{GM_\odot^2}{2a}$. For $a = 1.5\cdot 10^{13}$ cm, the kinetic energy of the stars is 10^{46} erg. If the distance between the two components is 1/100 of this figure (these binary systems are a fairly frequent occurrence), their energy is 10^{48} erg, i.e., comparable with the thermal energy of the stars. Thus, the closer the binary, the higher the mechanical energy of the two components. This conclusion is also readily seen to follow from the virial theorem (12).

Star systems which fill a large volume and comprise numerous stars, such as globular clusters in the Galaxy, contain a certain "dispersed" energy, apart from the kinetic energy of the individual stars and nebulae and the potential energy of their gravitational attraction. This "dispersed"

* Since magnetic fields exist on the surface of numerous stars, we should also estimate the magnetic energy of a star. This is a difficult problem, however, since the fields in the stellar interior are not known. The magnetic energy is probably much less than the thermal energy.

energy includes, in particular, the radiation of stars, cosmic rays, and magnetic energy distributed throughout the entire volume of the star system.

Each star in the Galaxy has a certain characteristic, so-called peculiar motion of its own relative to other stars. This velocity on the average reaches 10—20 km/sec. Hence, each star carries an additional kinetic energy of about $10^{45}-10^{46}$ erg. The velocities of stars associated with the rotation of the entire Galaxy around its center are greater than the peculiar velocities, reaching a few hundreds of kilometers per second. This gives another 10^{48} erg for each star in the Galaxy, and about $10^{59}-10^{60}$ erg for the entire system. According to the virial theorem, the potential energy of the Galaxy should be of the same order of magnitude.

The amount of "dispersed" energy in the Galaxy is thousands of times less than the mechanical energy, and it probably does not exceed 10^{56} erg. Since the density of this energy—i.e., the quantity of energy in unit volume volume — is very low, it does not convert on a large scale to other forms of energy, and its significance in connection with cosmic explosions is apparently not as prominent as that of the energy concentrated in compact celestial bodies.

§ 5. EXPLOSIONS ON THE SUN AND THEIR EFFECT ON THE EARTH

Because of our proximity to the Sun, even fairly weak phenomena on its surface, which mostly go by unnoticed on other stars, are observed and even felt on the Earth. These effects include among other things the fairly small (on cosmic scale) explosions on the Sun's surface, which are known as chromospheric flares. A tiny fraction of the energy released by a flare and reaching the Earth is sufficient to alter the state of the Earth's atmosphere, disrupt radio communications, excite polar aurorae, and other effebts. The study of chromospheric flares is therefore not only of theoretical interest. If we can penetrate into their real nature, we will be able to predict chromospheric flares and take measures in advance to minimize the harmful consequences of this solar phenomenon in our daily life.

Before going into a detailed discussion of the chromospheric flares, let us briefly consider some phenomena which take place in the outer layers of the Sun and are closely related to the flares. First we will describe the sunspot activity. Sunspots are one of the most prominent and best studied features on the surface of the Sun.

As we have already noted, the surface of any star, the Sun included, is that layer beyond which the observer cannot penetrate due to the opacity of the stellar matter. This surface and the adjoining layers of the Sun, which are known as the photosphere, are never completely quiescent. Observations through a telescope show that the photosphere is made up of bright cells or granules. These granules measure 500—1500 km across, and are separated from one another by darker streaks. The granules are short-lived formations. In a few minutes, the entire population of granules disappears and new granules develop in their place.

The granules are in fact jets of gas rising from the depths of the photosphere. This gas is approximately 100°K hotter than the surrounding medium, so that its density is less than the ambient density and it "floats" up to the surface of the Sun. At the surface the gas emits part of its energy into the space, cools down, grows heavier than the surrounding medium, and sinks back into the photospheric depths, yielding its place to new portions of hot gas. This mechanism of energy transfer from the solar interior is known as convective transfer of heat.

From time to time, numerous small dark areas develop between the granules, which merge into a single sunspot. The sunspots may reach a size of a few thousands and even tens of thousands of kilometers across. Sunspots mostly occur in groups, and each group shows two large and prominent "leading" spots. The area occupied by a group of spots often constitutes a considerable part of the entire solar disk. The lifetimes of the sunspots vary, ranging from a day or two to a few months.

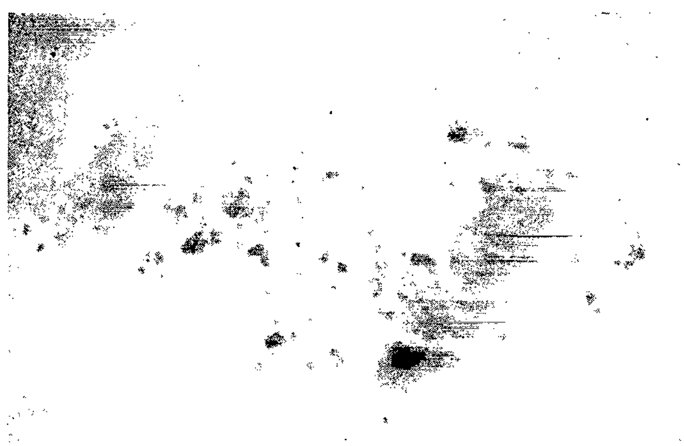

FIGURE 23. The structure of a large group of sunspots.

A group of sunspots moving across the disk will often disappear behind the Sun's limb and then reappear after about 14 days on the other side of the disk. This motion is apparently associated with the axial rotation of the Sun. Note that spots near the solar equator move across the disk faster than the spots far from the equator. Hence the conclusion that the Sun does not rotate as a rigid body does. Points on the equator complete one rotation around the axis in 25 days, whereas points at 30° latitude take 26.2 days. The reasons for this are not clear, but according to some theories this may be of significance in the sunspot activity and the related processes on the Sun.

The dark core of a sunspot is called the umbra (or the shadow). The lighter outer regions, constituting the penumbra, show a distinctly filamental structure. The temperature inside a sunspot is 1000—1500° lower than the temperature on the Sun's surface outside the spot. The emissivity of the gas in the sunspot is correspondingly lower. The emission per unit surface is proportional, as we know, to the fourth power of the temperature. Therefore, a difference of 25—30% in temperature between the sunspot and the photosphere reduces the spot brightness by a considerable factor compared with the photosphere. The sunspot appears dark only in contrast with the brighter photosphere, which is the background. In itself, the spot brightness is extremely high, as the temperature of the gas there reaches a few thousand degrees.

Since the surface of the Sun inside the sunspot emits less energy than outside the spot, certain factors apparently exist in the photosphere preventing free transfer of energy to the surface. The reduced emission from the sunspots is generally linked with the presence of strong magnetic fields.

A study of the Zeeman effect in the spectra of the outgoing radiation from sunspots shows that the magnetic fields in the spots reach several thousands of oersted.* The area of a sunspot may reach hundreds of millions of square kilometers. A field of this strength is set up at the center of a conducting coil of 5 cm radius if it carries a current of 1000 A. For comparison, note that the strength of the geomagnetic field acting on the compass needle is a few tenths of an oersted.

In the area occupied by a group of sunspots, the magnetic field is not limited to the spots as such: it overflows to the space between them, where it is weaker than inside the spots (reaching tens and sometimes hundreds of oersted) and highly inhomogeneous.

The polarization of radiation in the Zeeman components of a split line is indicatory of the direction of the magnetic field in the spots. The magnetic fields in the two major, leading spots are found to point in the opposite directions, in the sense that the North and the South magnetic poles of the Earth are oppositely magnetized. The magnetic lines of force inside the spots are approximately perpendicular to the surface of the Sun, i.e., the lines appear to emerge from one leading spot and sink into the other.

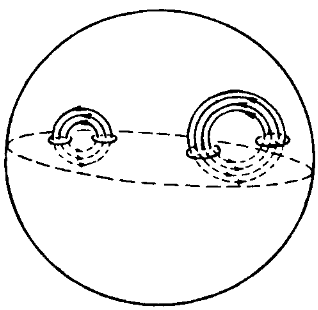

FIGURE 24. A diagram illustrating the relation between sunspots and magnetic fields.

It is a known fact that the magnetic field lines are always closed. The configuration of the magnetic lines between the two leading spots therefore suggests that a curved tube of field lines (a force tube) extends below the Sun's surface, joining the two leading spots in the same way that in the horseshoe magnet the curved iron bar joins the North and the South poles. Certain effects in the solar atmosphere seem to indicate that the magnetic lines of force form similar arches or loops above the surface as well. The sunspots thus form at points where the magnetic force rings cross the Sun's surface: part of the ring emerges above the surface, and the other part is sunk below the surface.

* A unit of magnetic field strength in the MKS system is ampere-turn per meter (A-turn/m). The relation between the CGSE and MKS units is $1\,\mathrm{Oe} = \frac{10^3}{4\pi}$ A-turn/m.

The strong magnetic fields under the sunspots also account for their lower temperature, compared to the rest of the Sun. As we have already mentioned, a considerable part of the energy radiated by the Sun is lifted by convection from deep-lying layers, i.e., by relatively hot gas masses rising to the surface. This gas contains ionized atoms, and it thus behaves as a plasma. The magnetic field affects the motion of charge particles: it actually inhibits the convective motion of the plasma under the sunspot. Therefore less energy reaches the surface in the sunspot than in any of the adjoining areas. As a result, the spot is colder and not as bright as the rest of the photosphere. Nearby areas around sunspots, on the other hand, occasionally become comparatively hotter, as is evident from the existence of the so-called faculae, which are relatively bright photospheric regions near sunspots.

Above the photosphere there is the solar atmosphere. This is a gas layer which absorbs radiation in discrete spectral lines and therefore emits at line frequencies only (we refer here to optical radiation). The general emission of the atmosphere is generally unnoticeable against the background of the bright photosphere, since the eye receives the light emitted by the photosphere in the entire optical range and the contribution from line emission is not large. The line emission is felt, however, during a total solar eclipse, when the photosphere is occulted by the disk of the Moon. The part of the atmosphere projected onto the dark sky forms a bright ring around the Sun. The spectrum of this ring consists of emission lines only. This is as it should be, since we have here a layer of hot gas which is transparent in the frequencies of the continuous spectrum. A relatively large proportion of the atmospheric emission is accounted for by the red hydrogen line H_α. The observed solar atmosphere thus appears red to the eye; this layer is known as the chromosphere (from the Greek "chroma," color). The chromosphere extends to altitudes of 15,000—20,000 km above the Sun's surface. Its structure is inhomogeneous. It appears made up of individual bright filaments, which extend all the way out, to the outermost rarefied reaches of the atmosphere, known as the solar corona.

Since the chromosphere envelopes the entire surface of the Sun, line emission is received not only from the parts of the atmosphere projected against the dark sky but actually from all the parts projected against the disk as well. As we have already noted, the genral emission of the photosphere "masks" the atmospheric line emission. At the line frequencies, however, the observed emission of the chromosphere is much stronger than the emission of the photosphere. Therefore, if the Sun's disk is observed in a very narrow spectral region, virtually inside a line, the instrument will receive mainly the emission of the chromosphere, and only a minor fraction of the light will originate in the photosphere. Observations of this kind are carried out with an instrument known as a spectroheliograph, which records the solar radiation in a very narrow range of wavelengths. Photographs of the Sun taken in the light of one of the strong spectral lines reveal the distribution of the emission in that line over the solar disk and are in fact photographs of the chromosphere.

The photographs of the Sun taken in the H_α hydrogen line (they are called spectroheliograms) reveal a number of interesting features. Different parts of the chromosphere show different emissivities: dark and light regions are scattered all over the disk. The bright regions, flocculi, are located near sunspots, and their H emission is stronger than the chromospheric average. Dark regions are extremely elongated and almost filamental in appearance.

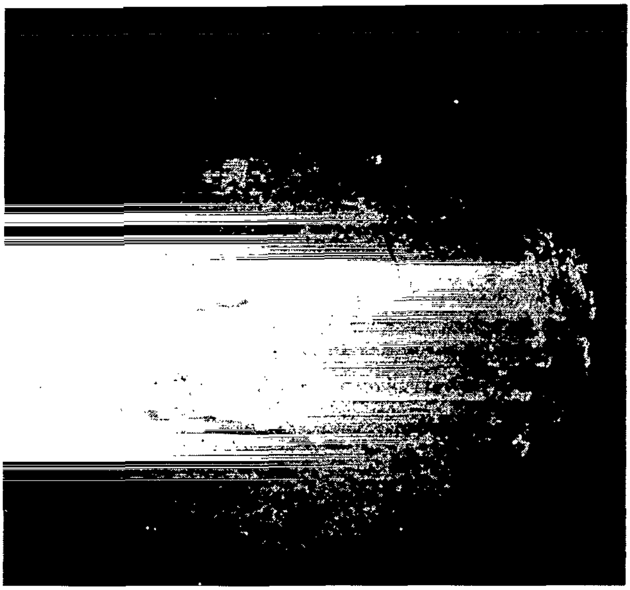

FIGURE 25. A spectroheliogram in the H_α light (4 April 1959).

The filaments appear dark only in contrast with the brighter adjoining areas of the solar disk. The rotation of the Sun will occasionally move a filament all the way out to the edge of the disk, where it is projected as an extremely bright formation against the background of dark sky. This phenomenon is known as a solar prominence. A dark filament is relatively long-lived: it persists up to a few months, slowly changing its form. Spectroheliograms also show other dark formations near sunspots, which are smaller than the filaments. At the edge of the disk they are observed as very short-lived eruptive prominences.

The motions in prominences were studied with the aid of motion picture films. The films clearly show that the eruptive prominences often move with tremendous velocities of several hundreds of kilometers per second away from the Sun's surface. In other cases, descending motions toward the surface are observed. Sometimes the prominences seem to pour into a sunspot.

The nature of the solar prominences is not clear to this day. These are apparently compressions of solar gas in the outermost reaches of the solar atmosphere, and they apparently are observed on the edge of the Sun owing to the enhanced emission of the upper atmosphere in the hydrogen lines. However, the origin of these compressions and their motions remain unaccounted for by the theory. In a number of cases the prominences rise in the form of an arch, apparently extending along the magnetic field lines from one sunspot to another. The motion of the prominences is often also guided by magnetic lines. Magnetic forces probably play a fundamental role in the formation and evolution of the solar prominences.

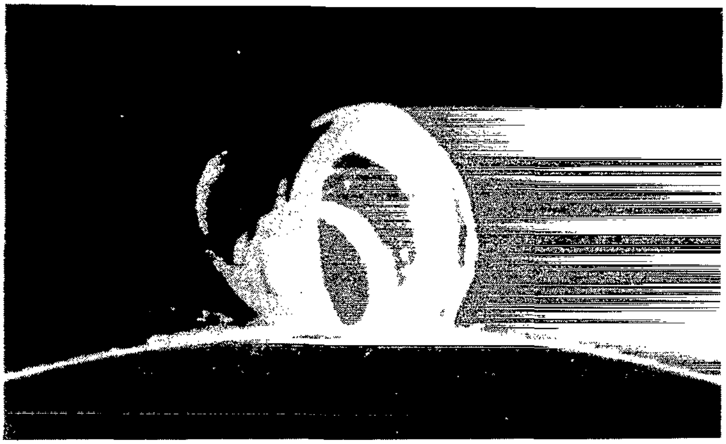

FIGURE 26. A solar prominence above a sunspot group.

Prominences and sunspots constitute what is known as the solar activity. The degree of solar activity varies periodically. In some years the total count of sunspots and their overall area are less than in other years. The time of maximum "spottiness" is known as the epoch of maximum solar activity. This epoch coincides with the maximum number of filaments and flocculi on the Sun. The cycle of solar activity, i.e., the time from one maximum to the next, is slightly more than 11 years. During each cycle, both the "spottiness" and the region of preferred sunspot formation change. After the minimum, at the beginning of each cycle, sunspots form fairly far from the equator, at latitudes of $30-35°$. Near the maximum, the average latitude of sunspot formation is $16°$.

The solar activity is a highly important factor influencing a wide range of terrestrial processes. Our environment is susceptible to various manifestations of solar activity, which affect, in particular, the state of the Earth's atmosphere and the life processes of certain organisms. We cannot describe all the different effects that solar activity has on terrestrial systems and organisms, and we will only briefly consider the outcomes of solar explosions.

Solar explosions are called chromospheric flares: they occur in the solar atmosphere. During the explosion, the brightness of a large area near a group of sunspots (generally the place previously occupied by a flocculus) increases abruptly. As a rule, this is associated with enhanced emission in hydrogen lines, but occasionally a very strong explosion affects the emission in the continuous spectrum as well. If this is so, the flare is visible in white light, against the photospheric background. Weaker flares are viewed through a spectroheliograph.

FIGURE 27. A large chromospheric flare observed in H_{α} light (18 September 1957).

The enhanced brightness in the flare area persists for a few minutes only. Then the brightness falls off and after 30—60 minutes the flare is no longer distinguishable from the rest of the solar disk. The area of enhanced brightness in a large flare may reach tens of thousands of kilometers. In the period of maximum solar activity, several strong flares may occur daily. In the epoch of minimum, the flares are fewer and invariably weak. In accordance with the regular migration of the sunspots from the equator to higher latitudes, the flares are observed near the equator in the period of maximum activity, and farther away at the minimum.

The flare is associated with turbulent motions of luminous gas with velocities of a few hundreds of kilometers per second. The gases move mainly upward, away from the Sun's surface, forming gas jets which rise to hundreds of thousands of kilometers. Although part of the gas in these jets will fall back to the Sun's surface, there are indications of continued emission of fast particles from the flare area into the outer space. Valuable evidence in this direction is supplied by the radio emission of the solar corona during flares.

The outermost layers of the Sun, known as the solar corona, extend over hundreds of thousands of kilometers above the surface; they emit very weak optical radiation, which is noticeable only when the photosphere is occulted (screened off) by the Moon or by an artificial screen. The solar corona is a gas in a highly peculiar state. Its temperature reaches 1—2 million degrees. Almost all the atoms in the corona are ionized, and it therefore contains a large number of free electrons. Because of its exceedingly high temperature, the solar corona is a source of radio waves; it is more powerful than the photosphere at the decimeter and meter wavelengths. The coronal radio emission is also observed between eclipses.

The propagation of radio waves in any medium is determined by the concentration of free electrons. The electron concentration of the corona was obtained from optical observations, and it was found to diminish with increasing distance from the photosphere. As the electron concentration falls off toward the periphery of the solar corona, the shorter radio waves originate closer to the photosphere, whereas the longer waves reach us only from the outermost layers. We cannot go into a detailed explanation of these effects, although it is of the greatest significance in the study of chromospheric flares.

A chromospheric flare is accompanied by a sudden enhancement of the radio emission from the part of the corona above the flare; this effect is known as a "radio burst." Bursts of different kinds are observed. The so-called type III bursts occur a few seconds after the beginning of the optical flare. Measurements of the burst level in the corona led to the conclusion that the factor responsible for the radio bursts propagates through the corona with a velocity of some 100,000 km/sec away from the Sun's surface. A few minutes after the burst, a more prolonged type II burst is observed, whose source propagates in the outward direction through the corona with a velocity of a few thousand kilometers per second.

Plasma oscillations are regarded as a possible source of radio bursts. If a volume of plasma locally acquires an excess of charged particles, say electrons relative to ions, all the charges in this volume start oscillating. Oscillating charge, as we know, produces electromagnetic radiation. These oscillations are set up, in particular, by passing a stream of fast charged particles through the plasma. The oscillation frequency is proportional to $\sqrt{n_e}$, where n_e is the electron concentration in the plasma. The electron concentration n_e of the corona is known, and the corresponding frequency is indeed close to the observed frequency. Thus, if the above assumption of the origin of the coronal radio waves is true, we conclude that fast charged particles are ejected from the solar flare. The group of particles creating a type III bursts move with velocities of the order of 100,000 km/sec through the corona, whereas type II bursts are apparently associated with the propagation of a shock wave through the chromosphere and the corona: this shock wave is associated with the expansion of gas in the flare, and it also produces plasma oscillations.

This conclusion derived from radio burst observations is consistent with the long-established observations of corpuscular ejection from flares. The existence of corpuscular streams is proved by numerous effects in the Earth's atmosphere which follow a solar flare; these effects primarily include the so-called magnetic storms. The state of the atmosphere is influenced by incoming particles of solar origin and by the electromagnetic radiation of the flare, which alters the electrical properties of the upper atmospheric layers. We will have to consider these properties in some detail before we can proceed with a discussion of the solar—terrestrial relations.

Long-range radio communications prove that radio waves propagate beyond the horizon, following the Earth's curvature. However, since radio waves, like light, are a form of electromagnetic radiation, they propagate in straight lines. The apparent contradiction between these two facts is only superficial. Radio waves do indeed propagate in straight lines, but the upper atmospheric layers at altitudes of some 100 km and higher reflect radio waves. Successive reflections from these atmospheric layers and from the Earth's surface are responsible for the long-range propagation of radio waves beyond the horizon.

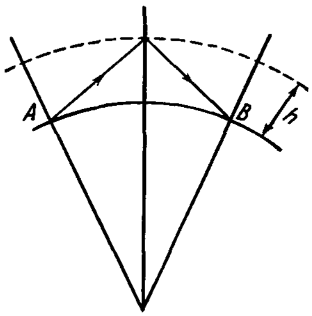

FIGURE 28. Reflection of radio waves from the ionosphere.

It is a known fact that radio waves are effectively reflected and scattered only by electric conductors. The reflection of radio waves from the upper atmosphere thus proves that these layers of the atmosphere conduct electricity. Near the ground the atmosphere is made up almost entirely of neutral atoms. At altitudes of 80—300 km, on the other hand, the atmosphere contains fairly high concentrations of ions and free electrons. These conducting layers of the atmosphere are called the ionosphere.

A wave emitted by a source on the ground is reflected, that is redirected back to the ground, following a gradual curving of the wave path in the ionosphere. The electron concentration in the ionosphere increases with altitude, which leads to an ever increasing refraction of the propagating radio wave and a corresponding change in the direction of wave propagation.

Before being reflected back to the ground, the wave should penetrate to a certain depth into the ionosphere. This involves a partial absorption of the wave energy. In principle, the absorption mechanism is very simple. The electrons in the atmosphere are made to oscillate by the electromagnetic field of the wave, and in their turn start emitting electromagnetic radiation. Part of the kinetic energy of the electrons, however, is transferred to heavy particles — ions and neutral atoms — which virtually do not emit any radiation. A certain fraction of the energy reaching the ionosphere is thus "turned off," as it is converted into other forms of energy. Thus, the lower the concentration of neutral atoms in the reflecting layer, the smaller the energy losses in reflection.

The reflectance of the atmosphere depends on the wavelength of the reflected radiation. Very short radio waves, e.g., those used in television, are not reflected, passing through the ionosphere unimpeded into the outer space. Television broadcasts therefore cannot be received beyond the horizon of the transmitting station. On the other hand, waves of about 10 m and longer are completely stopped by the ionosphere.* Radiation with wavelengths up to 30 m is reflected by the upper ionospheric layers, whereas longer waves penetrate only into the lowermost layers of the ionosphere.

The penetration of radio waves through the ionosphere or their reflection from this layer is determined by the concentration of free electrons in

* At night, waves of 10—30 m can penetrate through the ionosphere.

the ionosphere. The higher the number density of electrons per unit volume, the shorter the waves reflected by the layer.

Short radio waves (15—30 m) are best propagated at night-time. This fact is well known to any radio amateur, as well as another useful fact, namely that in winter radio reception is better than in summer. Since the range of propagation of radio waves is determined by the extent of their absorption in the ionosphere, we conclude that short waves are less absorbed by the ionosphere at night and in winter time. Since the change of day and night and the succession of seasons are entirely determined by the position of the Earth relative to the Sun, it seems that the variation in the propagation of radio waves is associated with the effect of the Sun on the state of the ionosphere.

How can the Sun affect the ionosphere? Apart from optical radiation, the Sun emits high-frequency photons, in particular in the ultraviolet and the X-ray regions of the spectrum. The energy of these photons is sufficient to dissociate oxygen and nitrogen molecules in the atmosphere into their constituent atoms and even to ionize the individual atoms. These ionizations produce the free electrons needed to make the ionosphere opaque for radio waves. The quantity of ultraviolet radiation and X-rays reaching the ionosphere thus determines the concentration of free electrons in the layer.

The dark side of the Earth is shielded from solar radiation, so that no atoms are ionized in night-time, whereas recombinations proceed unimpeded. The concentration of free electrons in the ionosphere thus diminishes at night. Radio waves reach progressively higher layers in the ionosphere, and the reflection energy losses are less than in the lower layers.

In winter the Sun is low above the horizon and its light, before penetrating sufficiently deep into the atmosphere, traverses a greater air mass than in summer, when the sunlight is almost perpendicular. The ultraviolet radiation in winter is absorbed in the upper atmsophere, and it reaches the lower ionospheric layers at 100—120 km altitude after having been considerably attenuated. The ionization at this level is thus lower than in summer and the radio waves can penetrate into higher parts of the atmosphere, where they are reflected with comparatively small energy losses.

Chromospheric flares produce a drastic change in the state of the ionosphere, which primarily affects the radio communications. The signal strength at short waves (10—30 m) is markedly attenuated. This "fadeout" effect is associated with increased absorption of radio waves. The waves are reflected from the lower layers of the ionosphere, where the concentration of free electrons has increased. A strong flare raises the electron concentration by as much as one order of magnitude compared to the concentration before the flare. We thus conclude that a chromospheric flare leads to an abrupt increase in the flux of high-energy photons from the Sun.

The change in the properties of the lower ionospheric layers during a chromospheric flare provides a method of flare detection which is independent of optical observations. Of particular interest is the enhancement of the very-long-wave radio signals (over 1 km long) generated by thunderstorm discharges. These signals, known as atmospherics, are generated continuously (and not only during thunderstorms) and produce characteristic noise and static in radio receivers. The ionization of the lower ionosphere following a chromospheric flare substantially improves the reflectance at these wavelengths, and the static therefore considerably increases during a flare: the atmospheric produced by distant discharges, which are normally not detected, come through strong and clear.

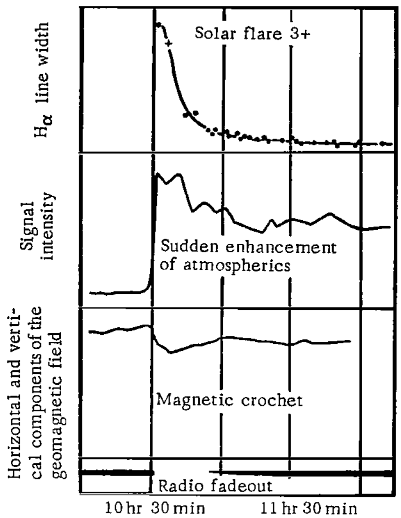

FIGURE 29. Ionospheric effects associated with a large solar flare (after Ellison).

Electric currents always circulate in the conducting ionosphere. They set up a magnetic field which is part of the general geomagnetic field. As the concentration of free electrons increases, the current is intensified, since the conductivity of the ionosphere increases and its resistance drastically falls. Chromospheric flares thus produce a change in the geomagnetic field, a so-called magnetic crochet (crochet is French for "hook"; this term reflects the shape of the curve which plots the field strength during a flare — see Figure 29).

The various ionospheric effects produced by a strong flare are shown in Figure 29. The top part plots the variation of the width of the hydrogen emission line H_α in the flare area. We have noted above that in the optical spectrum a flare produces an enhancement of the hydrogen line emission from certain part of the solar surface. This enhancement is associated with the formation of wide emission lines in the spectrum. The line width characterizes to a certain extent the strength of the flare, or alternatively the quantity of energy emitted in unit time.

We see that the various ionospheric events are triggered almost simultaneously with the beginning of the flare, gradually falling off in step with the decrease in flare brightness. This trend is associated with the enhancement of the ultraviolet emission of the Sun. Ionospheric observations thus prove that a flare produces an enhancement of ultraviolet and X-ray radiation, as well as optical light.

Let us estimate the energy emitted by a large flare in the H_α line. The area of a large flare is of the order of one thousandth of the solar disk. The H_α line occupies about 0.001 of the entire optical spectrum, where the solar disk emits a total of 10^{33} erg/sec. The flare area before the eruption thus emits in the part of the spectrum corresponding to the H_α line $10^{33} \cdot 10^{-3} \cdot 10^{-3} \approx 10^{27}$ erg/sec. During the eruption, the brightness in the H_α line is substantially higher than the brightness in the continuous spectrum. If the flare persists for 5 — 10 minutes (the phase of the maximum), it emits a total of 10^{30} erg in the line. This quantity in fact determines the total emission in the entire optical spectrum, since the emission in other lines is much less than in H_α. As regards the ultraviolet and the X-ray emission of the flare, ionospheric data and space probe findings indicate that their energy is several times, and possibly even one order of magnitude, greater than the optical energy. The total energy of photons ejected by a large chromospheric flare is thus of the order of 10^{31} erg. This corresponds to the energy released in the explosion of one billion of megaton bombs.

The ionospheric disturbances produced by the short-wave radiation of a flare occur simultaneously with the optical effects, since both types of radiation propagate with equal velocity. Magnetic storms and auroral enhancements on the other hand occur approximately after 24 hours. A magnetic storm constitutes a strong and prolonged (lasting for a day or two on the average) disturbance in the state of the atmosphere over the entire globe. The storm involves a breakdown in radio communications at all frequencies and may even interfere with telephone lines and other wire communications. This indicates that the upper layers of the atmosphere change their properties and partly break down. The ionization in the inner ionospheric layers markedly increases and the electric currents increase proportionately, so that the magnetic fields set up by these strong currents induce in their turn harmful parasitic currents in communication lines. The ionospheric currents also produce strong fluctuations in the observed geomagnetic field strength, and hence the name "magnetic storms."

The magnetic storms are generally linked with the action of charged particles — protons, electrons, and atomic nuclei — ejected from the chromospheric flare on the Earth's ionosphere. The particles accelerated by the shock wave (this is the wave responsible for the slowly propagating type II radio bursts in the solar corona) reach the Earth, where they are trappped by the geomagnetic field in the so-called radiation belts, which are distant a few thousands of kilometers from the Earth's surface. The energetic particles from the belts can penetrate into the ionosphere, destroying its top layers and producing ionization in excess of that caused by the ultraviolet radiation of the Sun. These disturbances are generally associated with flares which took place near the center of the solar disk. The corpuscular streams ejected by the flare and bombarding the Earth's atmosphere are thus directed approximately at right angles to the Sun's surface. These streams are clouds of plasma carrying their own magnetic fields. The field lines are "frozen in" and are thus transported with the plasma clouds.

The charged particles move along the magnetic lines of force at an altitude of a few thousand kilometers above the Earth's surface (in the radiation belts); near the magnetic poles the lines of force approach closer to the Earth's surface, and the charged particles therefore penetrate into deeper lying layers of the atmosphere. Colliding with atoms and molecules, the charged particles transfer part of their energy and thus excite the polar aurorae.

Particles which excite magnetic storms and polar aurorae should cover the distance from the Sun to the Earth (150,000,000 km) in approximately one day. They thus move with velocities of 1500—1600 km/sec, which approximately corresponds to the propagation velocity of type II radio bursts. This should be regarded as an average velocity: the magnetic storm has a finite duration, which means that not all the particles reach the Earth at the same time. On the other hand, all the particles seem to be ejected instantaneously either with the flare or soon after.*

Studies of the interaction of corpuscular streams with the geomagnetic field has shown that the mass of the entire plasma jet ejected during the flare is of the order of 10^{18} g. At velocities of 1500 km/sec, the kinetic energy of this stream is about 10^{34} erg. The energy released by a large chromospheric flare in the form of kinetic energy of ejected particles (corpuscular radiation) is thus thousands of times greater than the energy of optical radiation.**

Note that even in the absence of flares the Sun emits streams of corpuscules, which constitute the so-called "solar wind." However, the velocities of the solar wind corpuscules are a tenth of the velocities in corpuscular streams of flare origin.

Very strong flares produce still another remarkable effect on the Earth: they lead to an increase in the intensity of cosmic rays. Cosmic rays, which are described in more detail in a later section, are highly energetic particles (nuclei and electrons) moving with velocities close to the velocity of light. They fill the interstellar space. Particles of sufficiently high energies (hard radiation) cannot be effectively stopped by the weak geomagnetic field, and they readily penetrate at all latitudes to the ground. Less energetic particles (soft radiation), on the other hand, are deflected by the geomagnetic field and reach mainly the region near the magnetic poles of the Earth.

The enhancement of cosmic rays in high latitudes is noticeable about an hour after a strong chromospheric flare. Since no such effect is observed in the equatorial regions, it seems that only soft cosmic rays are generated by the flare. Seeing that these particles take 3—4 times as long to reach the Earth as the photons, we obtain an average velocity of about 100,000 km/sec for the cosmic rays. The energy of the protons (they constitute the bulk of the cosmic rays of solar origin) moving with these velocities is 10^{-3} erg per particle. Since each flare releases some 10^{32} erg in the form of cosmic rays, the total number of fast particles generated by a flare is about 10^{35}, which gives 10^{10} g of matter. These particles probably lose some of their energy while moving through the solar corona, where they create type III radio bursts.

The streams of cosmic rays and high-energy photons from chromospheric flares may prove hazardous to astronauts. These particles readily penetrate through a thick layer of matter, and the spacecraft hull does not provide adequate shielding. Therefore, for safety reasons, interstellar trips are best made in the period of minimum flare activity. The study of chromospheric flares and their origin may thus prove of primary importance for astronautics.

During a large chromospheric flare (500—1000 sec), a relatively small part of the chromosphere releases approximately the same amount of energy as the entire Sun during 1 sec. What is the source of this tremendous energy? There is still no conclusive answer to this question, but

* The jets of luminous gas moving up from the flare area have velocities which are 1/2—1/3 of the above figure. These are either different streams or the corpuscular stream is accelerated in its motion toward the Earth. Near the Sun, no streams with velocities of 1500 km/sec are observed.

** The fragmentary data on the X-ray emission of flares suggest that the X-ray energy, although much higher than the optical, is nevertheless lower than the energy of corpuscular streams.

as we shall see observational data seem to point to a certain connection of flares with magnetic fields.

Between sunspots of opposite polarity there is a point where the magnetic field vanishes; the magnetic field strength thus increases in either direction, and the field lines point in opposite directions. These points of zero magnetic field are called neutral points. Chromospheric flares are generally localized near neutral points, and they involve a change in field configuration near the flare, sometimes leading to a reduction in field strength. It is therefore generally assumed that the flare energy is derived from the magnetic energy of these fields. The exact process governing the conversion of magnetic energy into kinetic energy is not entirely clear, and we will therefore only estimate the amount of energy contained in the solar magnetic field.

A chromospheric flare encompasses the volume above an area of the order of $10^{19}-10^{20}\,cm^2$ (roughly 1/1000 of the disk area). The height of the chromosphere is about $2\cdot10^9\,cm$. This volume is therefore approximately $10^{29}\,cm^3$. The field strength in the active region near the sunspots, where the flares occur, may reach a few hundred of oersted. The energy of a magnetic field E_{mag} in unit volume — the magnetic energy density — is related to the field strength H by the equality

$$E_{mag}=\frac{H^2}{8\pi}. \qquad (19)$$

Using equation (19), we see that for $H = 100$ oersted, $E_{mag}\approx 4\cdot10^2\,erg$. The relevant volume thus contains $10^{32}-10^{33}$ erg of magnetic energy. If most of this energy can be very rapidly converted to radiant or kinetic energy, the result will be an explosion of the same strength as a large chromospheric flare. It nevertheless seems that the magnetic energy is not quite sufficient for sustaining large flares, especially, since initially the flare occupies a comparatively small volume. The final answer to this question of energy sources in chromospheric flares should apparently be sought in a general study of various cosmic explosions. Nevertheless, there is an obvious relationship between chromospheric flares and magnetic fields.

Even if it is proved that the magnetic fields play a decisive role in chromospheric flares, the exact nature of the flares will be fully understood only when we have elucidated the mechanism responsible for the generation of the solar field. No general theory is available to account for the origin of magnetic fields in the Sun and other cosmic objects. Some theories suggest that strong magnetic fields exist in the core of the Sun and part of their energy is periodically transferred to the outer layers. Conversely, one of the recent theories assumes a weak (of the order of 1 oersted) general magnetic field in the Sun, whose lines are localized in a thin layer near the surface. Differences in the rotational velocity of the Sun at different latitudes may periodically produce strong fields in the equatorial region. This theory adequately accounts for the peculiar features of the solar activity cycle, but it encounters considerable energetic difficulties. The field energy in the final account is the transformed kinetic energy of the axial rotation of the Sun, or more precisely the part of the kinetic energy which is determined by the greater angular velocity of the equatorial regions. This excess energy is sufficient to sustain magnetic fields for a few thousand years only. Thus, as it often happens in science, having devised a theory which adequately explains one group of phenomena, we are faced with the necessity of solving a new no less complicated problem, namely what sustains the differential rotation of the Sun at different latitudes.

§ 6. EXPLODING STARS

Stellar explosions range between wide limits in terms of their strength. We began with the weakest explosions — chromospheric flares, and will now continue with a discussion of explosions on a much grander scale — new stars or novae. An exploding star is primarily noted by an abrupt increase in its brightness, and these explosions are therefore often referred to as flares.

The power of the optical radiation of the largest of chromospheric flares does not reach 1/1000 of the total radiation power of the Sun. A relatively weak flare of this kind cannot be observed with present observational means on any of the average or bright stars, which are distant at least several light years from the Earth. And yet it seems that explosions analogous to chromospheric flares are characteristic of a great many, if not all, stars. To make such a flare observable, however, the star should emit in the optical spectrum hundreds of times less energy than the Sun emits. Therefore, the weakest stellar explosions, corresponding in their scale to chromospheric explosions, are visible only in the nearest dwarf stars, at distances not exceeding 30—50 light years (at greater distances the dwarfs are not visible altogether).

The faintest stars, as we know, are the red dwarfs. Their temperature is half the temperature of the Sun and their radius is smaller by one order of magnitude; they thus emit hundreds of times less energy than the Sun. In the optical region, the disparity between the radiation of a red dwarf and that of the Sun is even more striking, since the bulk of the energy emitted by a red dwarf is in the infrared part of the spectrum.

An M-type dwarf, known as UV in the constellation of Cetus,* was the first star on which weak explosions, corresponding to the large solar flares, were observed. This is one of the nearest stars: its distance from the Sun is 8.6 light years. Similar flaring red dwarfs are now known as UV Ceti stars.

A characteristic feature of the flares in UV Ceti star is that the brightness increases very rapidly at the beginning of each flare: no comparable fast rise is observed in other stars. The brightness may increase ten-fold in a few seconds. The flares differ considerably in their strength. In large flares, the total radiation of the star increases more than hundredfold, and in the weakest flares the increase is by a factor of few times, in no more than a minute or two. Then the radiation declines at a fairly gradual rate, and after 10—20 minutes the star is back to normal.

The optical radiation of a UV Ceti star in the normal state, between flares, has a power of some 10^{29} erg/sec. During a strong flare, the star

* A notation using one or two Latin letters is used for variable stars, whose brightness changes.

emits additionally over 10^{33} erg in the optical spectrum, and the energy emitted during a very weak flare is of the order of 10^{31} erg. Thus, flares of UV Ceti stars are comparable to the largest chromospheric flares of the Sun in terms of the energy radiated in the optical spectrum. Flares involving more than a ten-fold increase in the observed radiation from the star occur on the average once every few days, whereas very weak flares (which enhance the emission by a factor of 1.5—2) are much more frequent: tens of these flares may occur daily.

The likeness between the flares of UV Ceti stars and chromospheric flares is not restricted to equally fast increase of radiated energy (the initial rate of development of a chromospheric flare is possibly somewhat lower). Spectroscopic observations have shown that the hydrogen lines of a UV Ceti star are markedly broadened during the flare, and the energy emitted by the star at the corresponding wavelength increases. A similar effect is observed, as we know, in solar flares. This likeness notwithstanding, however, the spectra of the two kinds of flares show substantial differences. The emission of chromospheric flares as a rule markedly increases only at the line frequencies, whereas the major part of the energy radiated by a flaring UV Ceti star is in the continuous spectrum, mainly in the blue and the near-violet. This difference is possibly partly associated with the greater brightness of the solar photosphere in the optical region: against this bright background we can hardly expect to distinguish the enhanced continuum emission of the chromospheric flare.

Recently, powerful radio telescopes picked up radio waves emitted synchronously with the flares of some UV Ceti stars. These radio bursts are generally in phase with the growth of the optical radiation, although sometimes they preceded the optical flare by a minute or two. A radio burst generally persists for 10—15 minutes, and the radio energy radiated during this time is 1/100—1/1000 of the optical energy. The proportion of the radio frequencies in the total flare energy is higher than for the solar flares, where the radio burst energy is only 1/100,000 of the total emitted energy.

The reasons for the flares of UV Ceti stars and the energy sources converted into electromagnetic radiation are not clear at present. However, since on the whole the flares are very short and the brightness falls off back to normal very rapidly, it seems that the explosions occur in the outermost layers of the stellar atmosphere, where the gas density is low. Indeed, an explosion below the star surface would heat up the photosphere, which will then cool down only through emission of radiant energy into the interstellar space. Since the photosphere is opaque to radiation, the cooling process should take a fairly long time. A stellar atmosphere, on the other hand, is transparent to radiation in a wide range of wavelengths: thus it can be heated almost instantaneously and will then rapidly lose the extra energy by readily radiating it into outer space. In terms of their location, the flares of UV Ceti stars thus coincide with chromospheric solar flares.

The radio waves from chromospheric flares are generally related, as we have mentioned above, with the passage of streams of fast particles from the explosion site through the outermost layers of the solar atmosphere. The radio bursts of UV Ceti stars are apparently associated with a similar process. These radio bursts are so far the only evidence of corpuscular emission with velocities approaching the velocity of light during the flares. We observe only that part of the corpuscular energy which is eventually converted into optical and radio emission. A larger proportion of the flare energy, to judge from solar flares, is carried off by the fast corpuscular streams, in particular by cosmic rays, and also by the elusive X-rays and gamma quanta. Therefore, the value of 10^{33} erg cited for the optical energy emitted by a large flare of a UV Ceti star hardly characterizes the total flare energy. It seems that the overall quantity of energy released during such a flare is much greater, possibly of the order of $10^{34}-10^{35}$ erg.

Even if we assume the existence of strong local magnetic fields in the outermost layers of a UV Ceti star, it is highly improbable that energies of the order of $10^{33}-10^{35}$ erg can be contained in the stellar atmosphere. This led to a hypothesis according to which the energy was supplied in latent form by some source in the stellar interior. The energy is supposedly suddenly released in the interior, and it is mainly converted to kinetic energy — the energy of fast particles and photons, which then diffuse into the atmosphere. Unfortunately, in the frame of our present-day concepts of stellar energy sources we cannot point to any reasonable process in the stellar interior which will convert the energy into any other form, but kinetic. This stumbling block is only temporary, however, and we are confident that eventually it will provide a stimulus to development of new ideas and more complete understanding of stellar flares.

Flares which release tremendous quantities of optical energy (millions of times more than in the flares of UV Ceti stars) are observed in the so-called T Tauri stars and in some other objects. We will not discuss these explosions in any detail, since the information here is highly fragmentary and no comprehensive picture has emerged so far. Note, however, that the rates of ascent and descent of the light curve of T Tauri stars are completely unlike the solar flare curves. The flare in these stars goes on for a few hours and sometimes even for a few days. Whatever the exact nature of the flares of T Tauri stars, a substantial proportion of energy in these explosions is released in deeper-lying layers of the star, and not in the atmosphere, and thus heats the stellar surface. We should again emphasize that both T Tauri stars and other stars with luminosities higher than the red dwarfs may pass through weak explosions, but the resulting increase in radiation remains unnoticed against the background of the total radiation of the star.

The explosions of novae, or new stars, have been studied in the greatest details. A nova is any new bright star which suddenly appears in the sky where no other star has been previously visible. This star is naturally "new" only for the observer: the discovery of a nova does not reflect the actual birth of a new star, but it is rather a result of a sudden increase in the light of a faint star by a factor of tens and even hundreds of thousands. Gigantic cosmic explosions responsible for the appearance of novae are a fairly frequent phenomenon — several novae are discovered annually in the part of the Galaxy near the Sun. A nova exploding at a distance of tens of thousands of light years, say, probably remains invisible to the terrestrial observer, and it thus seems that the total number of novae exploding each year in the Galaxy is substantially greater than the number of recorded novae, probably over 100.

Since novae appear quite at random — we do not know in advance which star will explode — the very first stage of the explosion, which is naturally of the greatest interest for establishing the exact nature of the phenomenon, is often missed. The various hypotheses of nova explosions are therefore largely based on the results of direct observations of the consequences of these explosions.

The recorded novae, as we shall see in what follows, mostly occur at an average distance of a few thousand light years from the Sun. Optical observations are therefore still the only source of information about these phenomena. Neither radio waves nor other forms of radiation from novae have been observed so far, which is apparently due more to insufficiency of observational means rather than to total absence of radiation at other frequencies.

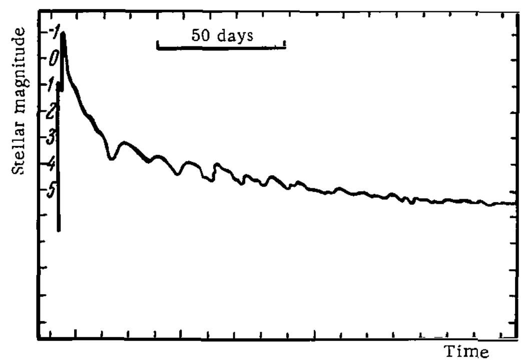

FIGURE 30. The light curve of a nova which appeared in the constellation of Aquila in 1918. The apparent magnitude at the maximum was only slightly less than that of Sirius, the brightest star in the sky.

The variation of stellar radiation with time is described by a so-called light curve. Figure 30 shows a typical light curve of a nova. The light radiation of the star (its brightness) is plotted in the figure in conventional astronomical units — stellar magnitudes. A stellar magnitude is defined as follows: an increase in the brightness of a star by a factor of 100 corresponds to a decrease of 5 in its stellar magnitude. A change of one stellar magnitude thus corresponds to a change by a factor of $\sqrt[5]{100} = 2.512$ in the brightness of the star. This system of units dates back to antiquity, and it is universally used in stellar astronomy, since a comparison of any two stars in stellar magnitudes is much easier than the comparison of their radiation in any system of energy units. The main advantage of this system for our purposes is that it permits drawing the light curves in more compact form.

We see from the light curve that the radiation emitted by a nova at first steeply increases to a maximum, and then gradually falls off with pronounced fluctuations.

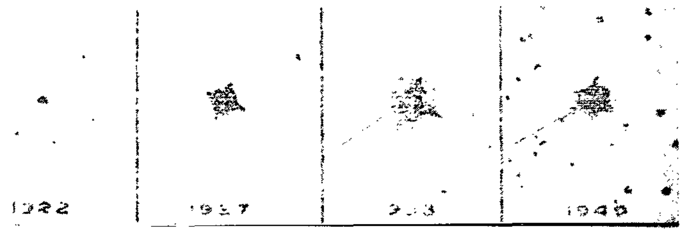

FIGURE 31. Photographs of Nova Aquilae 1918 taken in various years, clearly showing the expansion of the envelope. The photographs of 1933 and 1940 show a condensation of gas at the end of the envelope axis (marked with arrow).

A few years after the explosion, a bright nebula is often observed around the nova; these nebulae are regular circles or ellipses, not unlike a planetary nebula. The spectrum of these nebulae consists of emission lines, which is a sign of tenuous luminous gas. The apparent radius of the nebula, i.e., the angle that the nebula subtends in the sky for a terrestrial observer, increases with time. By measuring the rate of change of the apparent radius, we can establish the time when the nebula formed and the exact point where it originated. These measurements invariably show that the nebulae form at the time of the eruption of the nova, and the expansion center coincides with the star. Hence the conclusion that when a star explodes as a nova, its outermost layers are stripped off, forming an expanding envelope. After some time, when the envelope has expanded sufficiently, we can observe it as a nebula.

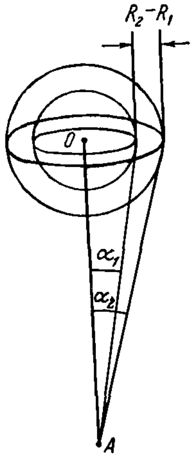

FIGURE 32. A schematic diagram illustrating the determination of distances of novae from observations of the expanding envelopes. In projection on the celestial sphere, the nebula appears as a ring.

The nebulae produced by nova eruptions appear as circles only in projection onto the sky. In fact these nebulae are spherical. The gas in the frontal part of the nebula moves toward the observer, and the gas in the rear half moves away from the observer. Doppler effect will thus shift the emission lines from the approaching part toward shorter waves and those from the receding part toward the longer red waves. The amount of shift is measured and gives the rate of expansion of the nebula. The nebulae produced by nova explosions are seen to expand with tremendous velocities, from a few hundreds km/sec to 1500—2000 km/sec. At these velocities, the radius of a nebula increases by $10^{16}-10^{17}$ cm in ten years ($3.2 \cdot 10^8$ sec).

Both the radius of the nebula and the rate of change of the angular size can be derived directly from a comparison of two photographs taken at an interval of a few years. If, say, the apparent radius of the nebula changes by the amount $\alpha_2-\alpha_1$, the change in the true radius R_2-R_1 is related to $\alpha_2-\alpha_1$ by the equation $R_2-R_1 = r(\alpha_2-\alpha_1)$, where r is the distance from the observer to the nova* (see Figure 32). Since $R_2 - R_1$ is measured independently of $\alpha_2 - \alpha_1$ from the rate of expansion of the nebula (which in its turn is determined from the line shift in the spectrum), this relation can be used to determine the distance r of the star. The difference $\alpha_2 - \alpha_1$ of the brightest novae reaches 10'' in 10 years. Since $R_2 - R_1 \approx \approx 10^{17}$ cm, these stars are thousands of light years away.

A highly important characteristic of the strength of the explosion is the amount of matter ejected into the interstellar space, i.e., in our case the mass of the expanding nebula. This mass is estimated from the known radius of the nebula, which is used to calculate its volume. The emission lines in the nebular spectrum, as we have mentioned in § 2, provide an indication of the relative content of the various elements and also the number of atoms of each element in 1 cm^3. Multiplying the number of atoms in 1 cm^3 by the mass of hydrogen atom (which is the most abundant element) and by the volume of the nebula, we obtain its mass. The masses are of the order of $10^{29} - 10^{30}$ g, which is approximately a hundred times the mass of the Earth.

In what follows we will discuss in more detail what stars explode as novae. We will see that these are dwarfs with radii one order of magnitude less than the radius of the Sun and masses a fraction of the solar mass. If the mass of the envelope is evenly distributed over the surface of such a star, we obtain for 1 cm^2 of the surface $10^{29}:(4\pi \cdot 10^{10})^2 \approx \approx 10^8$ g of matter. The mass of the stellar atmosphere per 1 cm^2, as we have seen above, is of the order of several grams. The ejected mass thus originated in the photosphere and in deeper-lying layers of the star. The gas pressure at the level of the explosion was equal to the weight of the stripped gas layers per 1 cm^2 of surface. This weight is $\frac{GM_*}{R_*^2} \frac{m}{4\pi R_*^2}$, where m is the mass of the envelope. Using the above values for M_*, R_*, and m, we find that the pressure at the level of the explosion was of the order of $10^{13} - 10^{14}$ g/cm $\cdot$ sec^2, i.e., millions of times the pressure in the stellar atmosphere. The corresponding temperatures were therefore very high — millions of degrees — and the gas density was also large. The layers above the explosion level naturally were completely opaque to radiation of all wavelengths.

Judging from the spherical shape of the expanding envelopes, we conclude that the explosion encompassed the entire layer of the star at a certain depth. The ejected envelope is therefore spherical, like the surface of the star, although the strength of the explosion is somewhat different in different directions, causing a certain "lopsidedness" of the nebula.

In the light of our discussion of nova explosions, we see that the eruption involves a change in the emission of the star. A tremendous amount of energy is released at a certain level in the star as a result of explosion. Part of this energy heats the gas in the exploding layer and the gas pressure increases to such an extent that it is no longer balanced by the weight of the overlying layers. The outer layers of the star therefore start expanding and the luminous surface increases, which is manifested in a rapid growth of the apparent brightness of the star.

* The angle $\alpha_2 - \alpha_1$ is found in radians, i.e., $\alpha_2 - \alpha_1 = \frac{(\alpha_2 - \alpha_1)''}{206\,265}$, where $(\alpha_2 - \alpha_1)''$ is the same angle expressed in seconds of arc.

The absorption lines in the spectrum of the nova are shifted relative to the normal positions toward shorter wavelengths. The reason for this violet line shift becomes clear if we remember that the eruption involves an expansion of the outermost layers of the star and that only the part approaching the observer is visible. The lines in the spectrum of the radiation from an approaching source are indeed shifted toward the violet according to the Doppler effect. Application of equation (5) shows that the surface of the nova moves with a velocity of the order of 1000 km/sec, which on the whole agrees with independent measurements of the expansion of envelopes.

The expanding outermost layers of the star form the envelope. As the radius of the envelope increases, the density per unit surface of course decreases. A few days after the eruption (i.e., after 10^5—$5 \cdot 10^5$ sec), the radius of the envelope of most stars will have exceeded 10^{13} sec, i.e., thousands of times greater than the radius of the star. On the other hand, the mass per 1 cm^2 of the envelope surface will have become millions of times less than the mass at the very initial stages, when the radius of the envelope was close to the radius of the star.* If initially the envelope contained 10^8 g of matter per 1 cm^2 of surface, the surface density will drop to about 10 g in a few days, i.e., a figure of the same order of magnitude as the gas density in the stellar atmosphere (per unit surface). The envelope will have become transparent to optical radiation in the continuous spectrum, just like the stellar atmosphere. The observer will not be able to see through the envelope.

Further increase of the surface of the expanding envelope beyond this point no longer increases the brightness. Prior to this stage, the envelope emitted at all the wavelengths of the continuous spectrum, but having become transparent it only absorbs at the frequencies corresponding to spectral lines, like the stellar atmosphere, and therefore emits only at the line frequencies. The observed brightness is mainly determined by the quantity of energy emitted in the continuous spectrum, i.e., the emission of the photosphere, while the envelope no longer acts as the photosphere. Hence, the peak brightness of the nova corresponds to the time when the envelope has just become transparent to optical continuous radiation.

Since we now "peep" into the expanding envelope, let us examine the processes which take place a few days after the gigantic explosion. The radius of the envelope after a few days is still small, so that no nebula is visible around the nova, and we can only study the spectrum of the radiation emerging from below the envelope. Since the emission of the envelope is superimposed on the stellar spectrum, we in fact observe a highly complex spectrum. It shows both continuum emission and lines (emission and absorption). The continuum emission is particularly strong in the blue and the violet, which is a result of the high temperature of the emitting gas.

Since the expanding envelope emits at line frequencies, the spectrum of the nova will naturally contain emission lines. When the envelope has become transparent, we can see its rear receding part, as well as the frontal part approaching the observer. The emission line from the entire envelope is therefore arranged symmetrically about the normal (unshifted)

* The mass per 1 cm^2 is inversely proportional to the square of the radius in an expanding nebula. This is evident from similar reasoning to that applied to equation (3).

position of the corresponding line, characterized by wavelength λ_0. The total width of the line is determined by the expansion velocity of the nebula v. The emission line is formed by radiation from different parts of the envelope, which move with different velocities relative to the observer. Since the maximum approach velocity is v, and the maximum recession velocity is $-v$, the total line width is clearly $2\Delta\lambda_0=2\lambda_0\frac{v}{c}$.

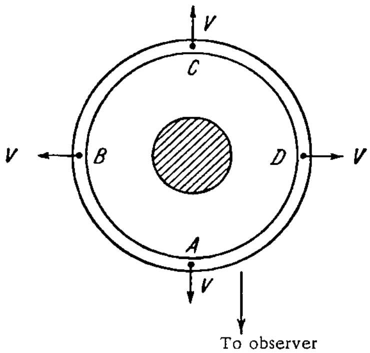

FIGURE 33. Schematic diagram of an expanding envelope. The part of the envelope between the star and the observer produces absorption lines shifted by the Doppler effect. Emission lines are produced by the entire envelope.

If a source of continuous radiation is located inside the envelope, the absorption of this radiation by the frontal part of the envelope moving toward the observer produces absorption lines with a shift $\Delta\lambda_0 = \lambda_0\frac{v}{c}$. These absorption lines are indeed observed in the spectrum. Because of the continuous radiation from the star, the brightness does not fall off steeply in the post-maximum stage, as could have been expected from the sudden reduction in the opacity of the envelope.

Apart from the lines of the stellar envelope, which are shifted by $\Delta\lambda_0$ due to the expansion, the post-maximum spectrum of a nova shows numerous other lines which are also displaced relative to thier normal position. The shift of these lines shows that they are associated with the absorption of photons from a source of continuous radiation by gas streams moving toward the observer a factor of 1.5—2 faster than the expanding envelope. Similar streams are seen to move in opposite directions. The study of lines in the spectra of novae has thus shown that after the explosion the star remains active for a few weeks or even months. It ejects masses of gas which are propelled with velocities of 2000—3000 km/sec. This gas both absorbs and emits at line frequencies, whereas the denser gas layers nearer the star also emit in the continuous spectrum. Detailed study of the line spectra of novae has established that the ejection of matter is distinctly asymmetric: the gas is ejected at different rates in different directions.

Although the ejection of gas from a nova gradually subsides, it may burst occasionally with renewed vigor. As a result, the emission in the continuous spectrum is intensified, and the brightness of the star increases. This effect is known as a secondary flare. Secondary flares are much weaker explosions than the first, primary eruption. Thus, the main explosion, which strips the outermost layers off the star, is followed by a succession of weaker minor explosions.

Since the velocities of the ejected gas streams are much higher than the velocity of the expanding envelope, the jets catch up with the envelope and accelerate it. The steady acceleration of the envelope is evident from the increasing shift of its absorption line. The most powerful jets are apparently ejected by novae still in the pre-maximum stage. Cutting into the envelope, they tear it up into separate fragments. Therefore the nebulae around ex-novae, which as we know are greatly expanded envelopes, are generally highly inhomogeneous, being made up of several patches or clouds.

The inhomogeneous structure of the nova envelope and its deviation from the spherical shape are at first not particularly noticeable. Gradually, however, as the nebula expands and the gas density diminishes, the gas emissivity falls off and the least dense parts of the nebula are no longer visible. At this stage, the nebula often appears made up of isolated luminous condensations, which are arranged not quite symmetrically about the parent star.

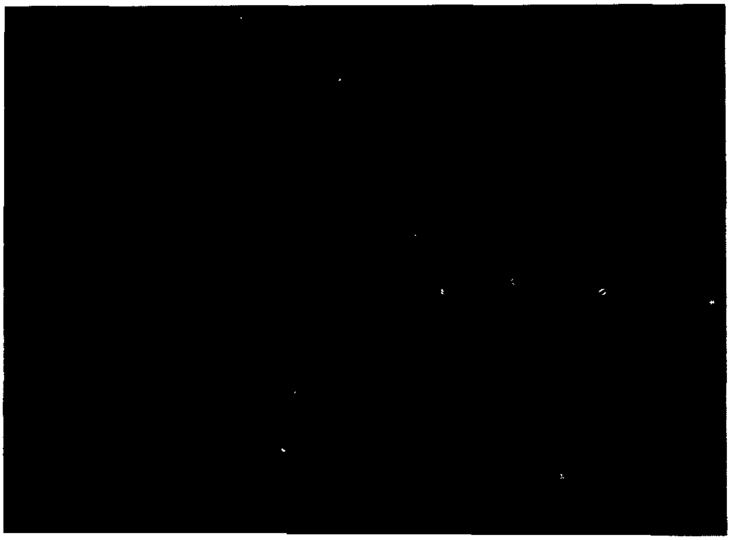

FIGURE 34. The envelope of Nova Persei 1901. The photograph taken 50 years after the eruption clearly shows the inhomogeneous structure of the envelope.

The structure of the expanding nebula was studied in maximum detail for one of the nearest novae, the one which flared up in the constellation of Aquila in 1918. The spectra of the nebula show that the bulk of the matter is concentrated in individual rings with a common axis. Two very prominent gas condensations are observed at the ends of the axis. One of these condensations showed on photographs for the first time in the 1940s, when it had receded sufficiently from the star (see Figure 31).

The flares of different stars follow different courses. The expansion velocities, the masses of the envelopes, and the characteristic features of brightness and spectrum variation are all different. All novae, however, acquire an expanding envelope as a result of the explosion and continue ejecting powerful gas streams for some time. In estimating the energy released by a nova we will therefore operate with some "average" star.

We start with the kinetic energy of the envelope. The mass ejected by the nova is estimated from the mass of the nebula around the star. The nebula may become visible a few months after the eruption, although usually a few years have to pass before the nebula can be resolved through a telescope. The ejection of gas from the star will have definitely stopped by then. Our previous mass estimates ($10^{29}-10^{30}$ g) therefore include the mass of the initially stripped layers and the mass of the subsequently ejected gas. Given the expansion velocity of $V \approx 1000$ km/sec $= 10^8$ cm/sec, we find for the kinetic energy of the nebula $E_{\text{kin}} = \frac{mV^2}{2} = 10^{45}-10^{46}$ erg.

The surface layers of a star are attracted by the deeper lying parts. To overcome this attraction, a certain energy is required. Let this stripping energy be E_{strip}. It can be estimated by the same method that we have used before in estimating the potential energy of a star (see § 4). This gives $E_{\text{strip}} = \frac{GM_* m}{R_*}$, where M_* and R_* are the mass and the radius of the nova, m is the mass of the envelope. Using the previous values for M_*, R_*, and m, we find E_{strip} of the order of 10^{45} erg.

Only part of the energy released by the explosion is effectively utilized in stripping off the envelope and accelerating it to the characteristic expansion velocities. The other part escapes in the form of radiation. The radiant energy of a nova, or more precisely the part of the energy in the visible spectrum, is estimated from the light curve.

The light curve plots the stellar magnitude as a function of time. Since the distance of the star is known, we can find the stellar magnitude that the Sun would have had at the same distance. Corresponding calculations show that at the time of maximum brightness an average nova is 12—15 magnitudes brighter than the Sun. Hence, its luminosity is 10^5-10^6 times that of the Sun, and the nova emits $10^{39}-10^{40}$ erg/sec as radiant energy. For the total quantity of energy emitted by the nova during the first few months after the explosion we find 10^{46} erg (making allowance for the decline of its brightness).

The radiation emitted by an erupting nova is not confined to the optical spectrum: it may also emit electromagnetic radiation at other wavelengths and even cosmic rays. Unfortunately, no data are available for estimating the energy of the invisible radiation from novae. There is, however, a possibility that their significance is smaller than in solar flares, since the explosion in a nova occurs deep in its interior and the energy of these components is largely converted to heat before diffusing to the surface. Anyhow, various estimates show that the total quantity of energy released by an exploding nova may exceed 10^{46} erg. In terms of energy, a nova is thus equivalent to $10^{12}-10^{13}$ large solar flares.

The light variation of the so-called recurrent novae is superficially reminiscent of the nova eruptions. These stars flare up every 30—40 years. In a few days the brightness increases by a factor of 10^3-10^4, and then falls off rapidly, so that a few months after the eruption it is almost back to normal. The spectrum of the star during the eruption shows wide emission lines. These lines are characteristic, as we know, of tenuous emitting gas, and their large width is probably associated with high velocities of motion.

The eruption of recurrent novae probably also produces an envelope, but the photographs of these stars taken after a few months or years show no trace of a nebula. The envelope of a recurrent nova apparently carries a comparatively small mass, $1/100-1/1000$ of the mass of an ordinary nova envelope. By the time the radius of the envelope reaches observable dimensions, the gas density and its emissivity will have dropped to such an extent that the light of the nebula does not form any image on the photographic plate. The kinetic energy of such a small envelope should be thousands of times less than the kinetic energy of a nova envelope. Estimates of the radiant energy emitted by a recurrent nova in the optical spectrum give $10^{42}-10^{43}$ erg.

A few years after the eruption, a recurrent nova is back to normal both in terms of its brightness and its spectrum. It thus seems that the eruption does not lead to a radical change in the state of the star, and the mechanism which caused the eruption will therefore eventually become operative again and cause still another flare.

Recurrent eruptions are also characteristic of the so-called nova-like stars, which are also called U Geminorum stars after their prototype. Major eruptions may occur on these stars several times a year. In each eruption, the brightness of the star increases in a few hours (and sometimes in a day or two) by a factor of $20-50$. The star retains its anomalous brightness for a few days, and then it rapidly returns to normal.

The spectra of these stars taken in between eruptions show very wide and strong emission lines associated with hydrogen atoms. The stars are thus usually surrounded by an envelope. The motions in this envelope, however, apparently do not involve orderly expansion. The flares do not affect the spectrum of the envelope: it remains virtually unchanged from one flare to the next. There are no indications either that the eruptions of U Geminorum stars lead to excessive ejection of stellar matter. The explosion probably encompasses only the outermost skin of the star. The total energy radiated in the optical spectrum by a flaring U Geminorum star is of the order of $10^{39}-10^{40}$ erg.

Until recently, superficial likeness of light variation was the only evidence in favor of the hypothesis which regarded the eruptions of novae, recurrent novae, and U Geminorum stars as essentially the same phenomenon manifesting itself on a widely different scale. However, observations of stars which erupted several decades ago as novae through the largest telescopes (the 200-in. reflector included) revealed them to be binaries. A similar fact was independently established for recurrent novae and U Geminorum stars.

In itself, the binary nature of these stars does not attract particular attention: after all, a very substantial percentage of the stars in the Galaxy are binaries. However, almost all the binaries which erupted in one time or another show distinct peculiarities: first, they are very close binary systems: the separation between the components is only several times their radius, and the orbital period of the stars about the common center of gravity is exceedingly small, of the order of a few hours; second, the components of these systems are invariably dwarfs: one of these is generally very small, not unlike a white dwarf. The masses of these stars are of the order of a few tenths of the solar mass. All this strongly suggests that the processes responsible for the eruption of these stars are not unrelated and probably have much in common.

Even before the discovery of the binary nature of U Geminorum stars and recurrent novae, an interesting feature was noted in the eruption of both types of these stars. First it was established that U Geminorum stars of longer cycle (i.e., longer period between successive eruptions) are characterized by stronger flares.* The same correlation was observed for the recurrent novae. The change in brightness during the eruption is stronger the longer the cycle. It is remarkable that this correlation was used in 1934 to predict the eruption of the recurrent nova T Coronae Borealis some 80—100 years after its previous eruption of 1866, and indeed the star erupted as a nova in 1946.

If we assume that novae erupt more than once in a lifetime and that this correlation is also applicable to them, we can predict the recurrence period of novae from their light variation during the eruption. The results give periods of a few thousand years!

In conclusion of this section we will briefly consider some hypotheses advanced as a possible explanation of nova eruptions. Any theory should start with the assumption concerning the sources of explosive energy. Some of the earlier hypotheses suggested that the energy is released as a result of a collision between two stars. The new observations, however, definitely disprove this hypothesis: only stars of a highly restricted class erupt as novae, whereas any two stars may collide. For this reason other theories which associate the eruption with random external interactions are also inadequate.

For some time the astrophysicists tended to attribute the eruption to a sudden collapse of the star, which led to an explosive conversion of potential energy into heat and radiation. The assumption of radical changes in the interior structure, however, is not fully applicable, definitely not for the recurrent novae. Moreover, the energy released in a gravitational collapse of a star would behave in an entirely different manner from what we observe in novae.

It seems to have been firmly established that, first, the eruption of a nova (and especially of a recurrent nova) is determined by interior processes taking place inside the star and, second, the eruption does not involve a drastic change in the structure of the star.

According to modern notions, the energy released in the eruption of a nova is supplied by nuclear reactions, but the opinions vary as to the exact location in the stellar interior where the responsible reactions take place. Some authorities suggest that the explosion occurs in the peripheral layers of the star, whereas others are of the opinion that the explosion energy is released deep in the central core. In the latter case it is assumed that the energy is transferred from the interior to the surface by a shock wave, which is quite capable of stripping up the outermost skin of the star and propelling it into space as an expanding envelope. This theory, however, is unable to explain a whole range of facts, such as the powerful ejection of gaseous matter from the star after the shedding of the envelope and the recurrent secondary flares. The school of thought that assigns the explosions in novae, recurrent novae, and U Geminorum stars to the peripheral layers thus appears to be closer to the truth.

None of the various theories of novae explains even the main observational facts. Future theories apparently must take into consideration the

* This is a statistical correlation, i.e., it is not necessarily true for every individual star, but on the whole it is valid for a large sample of novae.

binary nature of these stars, since the presence of a close neighbor may affect the interior structure and the energy releasing processes of the star. Moreover, strong magnetic fields on the surface may also play a certain role: their existence is suggested by the asymmetric ejection of gases from a nova after the eruption. The magnetic fields are possibly generated by the fast spin of the star in a close binary.

§ 7. SUPERNOVAE

Ancient chronicles and annals, side by side with historical events, recorded extraordinary celestial phenomena and apparitions. One of these extraordinary events, mentioned in different annals, was the appearance in 1054 of an unusually bright star, which remained visible for a whole month even in full daylight. It was noted that this star was tens of times brighter than Venus, which is normally the brightest object in the sky after the Sun and the Moon. The position of this extraordinary star, as reported in the ancient chronicles, is now adorned by a curious nebula, with the descriptive name of Crab Nebula. Comparison of photographs taken over a period of several decades established noticeable expansion of this nebula and made it possible to fix its age, which was found to be 900 years. It thus seems plausible that the Crab Nebula is associated with the explosion of the extraordinary star in 1054.

The distance of the Crab Nebula was found as for the nebulae around novae from a comparison of the rate of change of its angular diameter with the absolute expansion velocities measured in terms of the line shifts in the spectrum (see § 6). This distance was found to be 5000 light years, i.e., greater than the distance of many of the recent novae. And yet, the brightness of the star during the eruption was hundreds of times the brightness of the brightest novae. Hence, the Crab Nebula formed in a much more powerful explosion than the usual eruptions of novae. These exploding stars are classified as supernovae.

Supernova explosions are cosmic catastrophes on a gigantic scale. They are visible even when the star is located in a distant stellar system. For example, a supernova explosion was recorded in 1885 in a galaxy known as Andromeda Nebula, which is distant millions of light years from our star system. This star was thousands of times brighter than the ordinary novae observed in this galaxy. A sudden increase in the brightness of some distant galaxy is a sure indication of a supernova, since the brightness of one such explosion is comparable with the total brightness of billions of normal stars in the stellar system.

Supernovae are an exceedingly rare phenomenon in our Galaxy. Among the various stellar explosions noticed and recorded in the last two thousand years, there were less than 10 certain supernovae. The most recent ones were observed in 1572 by Tycho Brahe and in 1604 by Kepler. Any other supernovae which possibly exploded in the Galaxy in subsequent centuries remained unnoticed, probably because of strong absorption by interstellar dust.

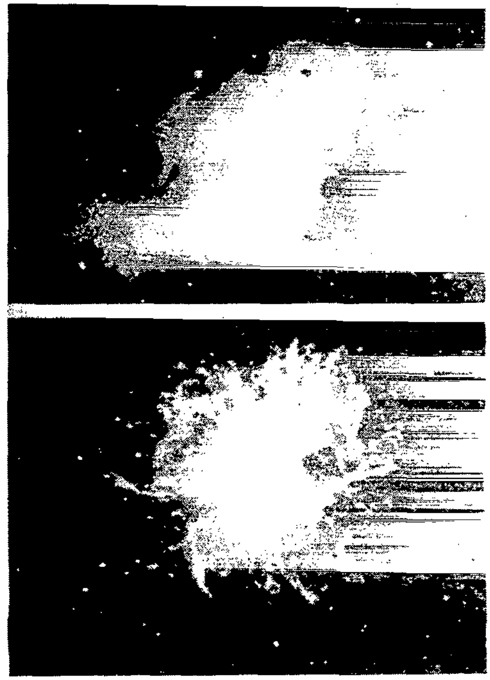

FIGURE 35. Crab Nebula. Top: photograph taken in the continuous light; bottom: photograph in the light of the H_α line.

All that we know about the light and spectrum variation of an exploding supernova is based on observations of stars in other galaxies. In our Galaxy we can only study the relics of supernovae, those nebulae which formed hundreds and thousands of years ago during supernova explosions.

A supernova in another galaxy is visible only during the very first stages of the eruption, when its brightness is sufficiently high to be distinguishable against the general mass of stars in the galaxy. A systematic search for supernovae in other galaxies is being regularly conducted for the last thirty years by periodically photographing large groups of galaxies. If a routine comparison of the photographs of the same part of the sky reveals a sudden increase in the brightness of one of the galaxies (this is evident from a stronger blackening of the negative), special observations of the light variation of that galaxy are instituted and the light curve of the supernova is obtained in this way. These crude observations clearly cannot capture the true beginning of the eruption, i.e., the ascending branch of the light curve, nor the true peak itself.

A supernova in another galaxy is bright enough to remain visible for a few months. During this time the light curve is plotted and the spectrum is photographed. Prior to 1965, over a hundred supernova explosions were discovered, but only few of them were covered by more or less detailed observations. Note that a comparison of the number of recorded supernovae with the total number of galaxies photographed in the systematic survey gives an estimate of the frequency of the supernovae. Supernovae are found to erupt once every 400—600 years in an average galaxy, but in giant spirals, like our Galaxy, the frequency of occurrence is several times higher.

The light variation of a supernova is on the whole similar to that of a nova, though on a much grander scale. The very fast rise (the maximum is reached in a few days) is followed by a comparatively slow decline. The shape of the descending branch clearly distinguishes between two types of supernovae, which are also markedly different in many other respects, as we shall see.

In supernovae of type I, the brightness decreases monotonically, without any oscillations, and the descending branch stretches over a fairly long time. During one year after the explosion, the brightness falls by a factor of several hundreds. All supernovae of this type have virtually identical light curves, whereas ordinary novae greatly differ in the fine features of their light variation. Supernovae of type II are characterized by a greater variety of light curves and by a rapid fall of brightness about 100 days after the maximum.

As we know, the energy radiated by a nova during the eruption is estimated from the light curve (if the distance of the nova is known). The energy radiated by a supernova in the optical spectrum can be estimated in the same way. For some of the galactic supernovae, maximum brightness estimates are available from ancient chronicles, and for the supernovae of 1572 and 1604 accurate light variation estimates were made by very experienced astronomers.

The distances of these supernovae were recently determined from measurements of the relic nebulae. The distances of supernovae exploding in other galaxies are found by special methods, which are discussed in the next section.

At the maximum a supernova radiates hundreds of millions of times the amount of energy radiated by the Sun. Type I supernovae are somewhat brighter than type II supernovae, emitting at a rate of 10^{42} erg/sec, whereas a type II supernova emits $10^{40}-10^{41}$ erg/sec. During the entire period of visibility, a type I supernova emits 10^{48} erg of optical energy, and a type II supernova about 10^{47} erg. The radiation of the supernovae clearly is not restricted to the optical spectrum. Possibly the total quantity of energy is 10—100 times the above figures, reaching $10^{49}-10^{50}$ erg.

The spectra of novae provided fairly detailed information on the changes which occur at the time of the maximum and soon after. The spectra of supernovae, on the other hand, are inadequate for reconstructing the entire picture of the explosion. The situation is somewhat better as far as type II supernovae are concerned, and we start with the description of this type.

The spectra of type II supernovae photographed after the maximum are on the whole similar to the spectra of novae in the corresponding phase. The most prominent feature, as in the spectra of novae, are the emission lines, which are wide enough to be called bands. If the emission in these bands is associated with an expanding envelope, and the band width is related to the expansion velocity by the Doppler effect, very high velocities of expansion are obtained. The standard Doppler effect equation is $\frac{2\Delta\lambda_0}{\lambda_0} = \frac{2v}{c}$ (see § 6), where $2\Delta\lambda_0$ is the band width, v is the wavelength of the corresponding line without Doppler shift, and λ_0 is the expansion velocity of the envelope. Observations give $\frac{2\Delta\lambda_0}{\lambda_0} \approx 0.04$, so that $v \approx 6000$ km/sec. The envelopes of novae never show such tremendous velocities.

Because of the very large band widths, the lines of different elements overlap and the spectrum of a supernova is completely unlike the spectrum of a normal star. The exact identification of the various bands is therefore highly uncertain, and the corresponding estimates of the expansion velocity are not very accurate. However, as no other data are available, we have to use in the first approximation the expansion velocities obtained from band width measurements.

The spectra of type I supernovae, on the other hand, remain completely puzzling to this day. Neither lines nor bands observed in the spectra of other stars can be identified here. The only exception are two lines in the red region of the spectrum, whose wavelengths coincide with the two well-known lines of oxygen. They appear only a few months after the maximum, and their width corresponds to velocities of the order of 1000 km/sec. In other respects, the spectrum of a type I supernova consists of a number of high intensity intervals alternating with darker gaps. The width of these intervals and their position are variable. It is significant that these changes occur always at the same phase in all type I supernovae.

The spectra of supernovae still do not provide adequate information on the nature of these explosions. This explains the great importance attached to the observations of nebulae produced by supernova explosions. The explosion leading to a supernova eruption is so powerful that the stellar envelope is much more massive than the envelope around a nova, and it remains visible for hundreds and thousands of years. This is evident from the very existence of the Crab Nebula, which is a remnant of the Supernova of 1054 that exploded in the constellation of Taurus. About a dozen bright nebulae are already known, whose origin is definitely associated with supernova explosions. Besides the fairly near Crab Nebula, the most remarkable specimens include the filamental nebulae in the constellation of Cassiopeia and in Cygnus.

The nebula produced by a supernova explosion in another galaxy is much less luminous than the entire star system and its size is vanishingly small compared to the size of the galaxy. These nebulae are therefore invisible, and we can only study the supernova relics in our Galaxy.

In the 1950s the Crab Nebula was identified as a powerful source of radio waves. It was later shown that all the other supernova remnants also emit large quantities of energy in the radio spectrum. The optical emission of these nebulae is often much weaker than their radio flux.

Observations of the radio emission of supernova remnants provided a highly powerful tool for unraveling the physical processes in these objects and thus advanced us a large step forward toward understanding the nature of supernova explosions. To understand the nature of the radio emission from the Crab Nebula and other related nebulae, we shall have to consider in some detail the various mechanisms which generate electromagnetic radiation.

FIGURE 36. The Loop in Cygnus (a system of filamental nebulae).

We have mentioned in the preceding (§ 2) that electromagnetic radiation is generated when electrically charged particles accelerate or decelerate. In a plasma, the thermal velocities of the individual particles are constantly changing because of collisions with one another. The radiation emitted in the process is known as thermal radiation. Another mechanism of electromagnetic radiation emission by plasma oscillations was mentioned above as a possible reason for solar radio bursts. However, the observed distribution of radio frequencies in the radiation from supernova remnants is entirely different from the predicted spectrum of these thermal mechanisms. Furthermore, optical radiation is also generated in both cases, and under the characteristic conditions of the nebulae it should be much stronger than the radio waves. The astrophysicists therefore had to look for other emission mechanisms which could account for the radio waves from the Crab Nebula and other related nebulae. One of the most promising candidates is the so-called gyromagnetic radiation or magnetic bremsstrahlung, which is now considered in some detail.

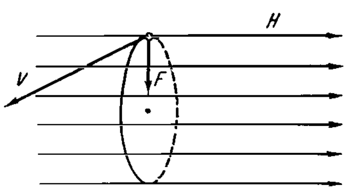

FIGURE 37. The direction of the Lorentz force F acting on an electron in a magnetic field. The direction of the electron velocity v and the field H is shown by arrows.

As we know, an electron moving in a magnetic field transversally to the field lines experiences a so-called Lorentz force which curves its trajectory. If the velocity of motion v is perpendicular to the lines of force, the Lorentz force F is given by the relation

$$F = \frac{e}{c} vH, \tag{20}$$

where H is the field strength and e is the electron charge. The force F is at right angles to the electron velocity and the field lines (Figure 37).*

If the magnetic field is constant (homogeneous) along the electron path, the force F will guide the electron in a circle. The radius of this circle R can be found by taking the Lorentz force F equal to the balancing centrifugal force $\frac{m_e v^2}{R}$ (m_e is the electron mass). Using equation (2), we find $R = \frac{m_e vc}{eH}$.

The motion of an electron in a homogeneous magnetic field is thus periodic, with a period equal to the time required to complete one circuit. This period is given by $\frac{2\pi R}{v} = \frac{2\pi m_e c}{eH}$. For the frequency of this motion we find

$$\nu_H = \frac{eH}{2\pi m_e c}. \tag{21}$$

This is also the frequency of the electromagnetic oscillations produced by the gyrating electron. The radiation emitted by an electron in a magnetic field is thus of frequency ν_s. There is also emission of radiation with frequencies which are multiples of ν_s, because the electron motion is periodic with all the multiple frequencies, as well as with the fundamental frequency ν_s. The multiple frequencies $n\nu_s$ are known as the overtones, and the number n is the order of the overtone. The period of an n-th order overtone is clearly $1/n$ of the fundamental period of the electron motion. For v much less than c, the bulk of energy is radiated at the fundamental frequency ν_s, and not at the overtones. Note that the expression of the fundamental frequency (21) is on the whole applicable only when the velocity of the electron is small compared to the velocity of light; otherwise, certain corrections should be introduced in this expression.

Let us estimate the frequency of the gyromagnetic radiation in a weak magnetic field, $H = 10^{-3}$ oersted, say. Inserting this H in equation (21) and setting $e = 4.8 \cdot 10^{-10}$ e.s.u., $m_e = 9.1 \cdot 10^{-28}$ g, we find $\nu_s = 2.8 \cdot 10^3$ sec^{-1}, and the corresponding wavelength is tremendous, $\lambda_0 \approx 10^7$ cm. Thus, gyromagnetic radiation in centimeter and decimeter waves can be observed only in sufficiently strong fields with $H = 10^2 - 10^3$ oersted. It is not quite probable that such very strong fields fill the entire tremendous volume of the nebula. The gyromagnetic mechanism in its classical form is therefore inapplicable to explaining the origin of radio waves in supernova envelopes.

The expression for the gyromagnetic radiation is different if the electron moves in a magnetic field with a velocity close to the velocity of light. The energy of this relativistic electron is much greater than the energy corresponding to the rest mass m_e. Relativistic electrons in space, together

* If the velocity is not perpendicular to the field, the expression for the Lorentz force contains the projection of the velocity on a direction at right angles to the lines of force. Then the electron trajectory is a combination of two motions: circular motion around the field lines and translation along the lines, so that the resultant trajectory is a helix.

with other high-energy particles, are a component of cosmic rays, which are generated, in particular, by cosmic explosions. Particles can be accelerated to velocities close to c in special laboratory installations. It was noted that electrons moving with velocities close to c in a strong magnetic field in a synchrotron accelerator start emitting light. Hence the term synchrotron radiation for the electromagnetic radiation of relativistic electrons moving in a magnetic field. Synchrotron radiation is fundamentally similar to the gyromagnetic radiation considered in the preceding, but the relativistic effects greatly modify the properties of this radiation.

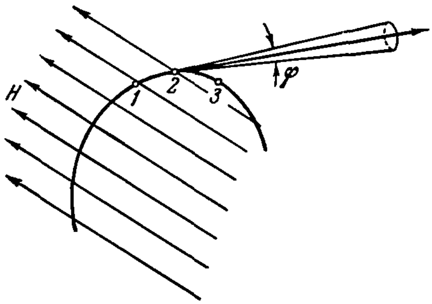

FIGURE 38. The radiation of a relativistic electron in a magnetic field. The plane of motion of the electron is perpendicular to the magnetic lines of force.

Classical gyromagnetic radiation is emitted isotropically in all directions, at right angles to the field lines. Synchrotron radiation, on the other hand, is emitted in a narrow-beam *in the direction of the instantaneous* motion of the electron. We will try to explain what causes this peculiar effect.

Consider the emission of a relativistic electron moving along the arc of a circle of length Δs between the points $1-2-3$. The velocity of the electron at the point 2 is toward the observer, and the points 1 and 3 are symmetric relative to point 2; the entire arc is so small that it can be treated as a segment of a straight line (Figure 38). Let the electron traverse this distance in a time $\Delta t'$. The radiation emitted by the electron will have propagated in this time to a distance $\frac{1}{2}(c\,\Delta t' - \Delta s)$ to the left of point 2 and to a distance $\frac{1}{2}(c\,\Delta t' + \Delta s)$ to the right of point two, toward the observer. Since the electron velocity is very close to c, Δs is not much different from $c\,\Delta t'$. Therefore, the volume of space filled by electromagnetic radiation to the right of the electron is much greater than to the left, and the energy is thus mainly concentrated to the right, toward the observer. In other words, the moving electron mainly radiates in the direction of its motion. Note that the electron does not radiate at right angles to the plane of its motion, since it is not accelerated in this direction.

The theory of motion of a relativistic electron in a magnetic field shows that an electron moving with a velocity v radiates in a very narrow cone, with an opening angle φ given by

$$\varphi \approx \sqrt{1 - \frac{v^2}{c^2}}, \tag{22}$$

and the axis of this cone is along the instantaneous velocity vector of the electron.

Synchrotron radiation was identified as one of the principal fundamentally novel features of the remnants of cosmic explosions. We will therefore derive an expression which describes the frequency of synchrotron radiation as a function of the electron energy.

The frequency ν_0 of a relativistic electron in a magnetic field is described by an expression analogous to (21), where the mass is no longer constant: the mass is a function of the electron velocity. This dependence is determined by equation (15), and for ν_0 we thus have

$$\nu_0 = \frac{eH}{2\pi m_e c}\sqrt{1-\frac{v^2}{c^2}}. \tag{23}$$

Since $1-\frac{v^2}{c^2}$ is much less than unity, the frequency of a relativistic electron is small compared to the corresponding frequency for nonrelativistic electrons.

A relativistic electron does not radiate at the frequency ν_0. Its radiation can be represented as a succession of very brief flashes. The electron completes one revolution in a time $\frac{1}{\nu_0}$, and emits in a given direction only for a fraction $\frac{\varphi}{2\pi}$ of this time, which we designate as $\overline{\Delta t}$. From (22) and (23) we find $\overline{\Delta t} = \frac{m_e c}{eH}$. If the observer were to move with the same velocity as the radiating electron, i.e., if the observer were at rest relative to the source of radiation, the observed duration of the flash would also be $\overline{\Delta t}$. The electron, however, moves toward the observer with a velocity v, and the observer therefore sees the flash for a shorter length of time $\Delta t = \overline{\Delta t}\left(1-\frac{v}{c}\right)$. This effect is fundamentally similar to the Doppler shift of electromagnetic radiation. For v very close to c we have

$$\Delta t = \frac{m_e c}{eH}\left(1-\frac{v}{c}\right) = \frac{m_e c}{eH}\,\frac{1-\frac{v^2}{c^2}}{1+\frac{v}{c}} \approx \frac{1}{2}\,\frac{m_e c}{eH}\left(1-\frac{v^2}{c^2}\right). \tag{24}$$

Since the duration of the flash is only a tiny fraction of the revolution period of the electron, the radiation in the flash corresponds to some high-order overtone of the frequency ν_0. The frequency of this overtone ν_c is of the order of $\frac{1}{\Delta t}$, i.e.,

$$\nu_c = \frac{2eH}{m_e c}\,\frac{1}{1-\frac{v^2}{c^2}}. \tag{25}$$

A relativistic electron also radiates at an infinity of other overtones, i.e., it emits radiation in a wide band of frequencies, which is essentially a continuous spectrum. The maximum radiation, however, is concentrated at the frequency $\bar{\nu} \approx 0.5\,\nu_c$, and the intensity falls off steeply on either side of this frequency.

The energy of a relativistic electron is described by the expression

$$E = \frac{m_e c^2}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{26}$$

which is derived using equation (17) for the relation between mass and velocity and remembering the equivalence of mass and energy (see § 4).

Using equations (25) and (26), $\bar{\nu}$ can be expressed in the form

$$\bar{\nu} \approx \frac{eH}{m_e c}\left(\frac{E}{m_e c^2}\right)^2. \tag{27}$$

Since the energy E of a relativistic electron is substantially greater than $m_e c^2$, it will emit photons of much higher energy than a nonrelativistic electron. Let $H = 10^{-3}$ oersted, and the electron velocity v differ from c by 0.01% (the energy E of the electron in this case is $10^4 m_e c^2$). The maximum radiation intensity is emitted by these electrons at a frequency $\bar{\nu} \approx 1.8 \cdot 10^{12}\,\mathrm{sec}^{-1}$. The wavelength of this radiation is about 10^{-2} cm, which lies in the infrared part of the spectrum, whereas a nonrelativistic electron in these fields emits very long radio waves.

Equation (27) shows that the frequency $\bar{\nu}$ at which the electron mainly radiates depends on the energy of the electron. Any real source of synchrotron radiation contains electrons with a certain energy distribution, which is known as the energy spectrum of the electrons. The synchrotron radiation from this source therefore occupies a wide frequency band in the continuous spectrum, and the energy distribution of the radiation is determined by the energy spectrum of the source electrons.

A relativistic electron emits in one plane perpendicular to the magnetic field lines. The electric field vector of the electromagnetic wave radiated by the electron lies in the same plane. The synchrotron radiation is therefore always linearly polarized, and the plane of polarization is perpendicular to the magnetic field. The polarization of synchrotron radiation is its best identifying characteristic, and it enables us to isolate it from the unpolarized thermal radiation. If the continuous radiation from a certain source displays a substantial degree of linear polarization, we can safely identify the synchrotron component of this radiation and even determine the direction of the magnetic field in the source.

Synchrotron radiation is now the only mechanism which can explain the radio emission of the Crab Nebula and other supernova remnants. A weighty argument in favor of this interpretation was its successful application to the optical radiation of the Crab Nebula. Its synchrotron origin was brilliantly confirmed when the light of the nebula was found to show considerable polarization. The final proof was obtained when the direction of the magnetic field derived from the polarization of optical radiation was found to coincide with the direction of the magnetic field derived from radio polarization observations. Since the frequency of optical radiation is tens of thousands times higher than the radio frequency, we see from equation (27) that the energy of electrons responsible for the emission of optical photons should be hundreds of times the energy of electrons radiating in the radio spectrum. Hence, the fairly strong optical radiation of the Crab Nebula is an indication of electrons of very high energies. In this respect it is markedly different from other supernova remnants, which are optically weak nebulae virtually without any synchrotron component in the optical spectrum. These nebulae apparently do not contain ultrarelativistic electrons.

Observations of the radio and optical spectrum of a nebula give the number of electrons of various energy. For the Crab Nebula the number of electrons of energy E was found to be proportional to $\frac{1}{E^{1.6}}$ for relatively low-energy electrons (those emitting in the radio spectrum) and proportional to $\frac{1}{E^4}$ for electrons emitting in the optical spectrum. Using $H \approx \approx 10^{-4}$ oersted for the probable magnetic field in the nebula (this figure was obtained from radio flux observations), the total energy of the relativistic electrons can be determined. The Crab Nebula was found to contain about 10^{47} relativistic electrons with a total energy of the order of 10^{48} erg. The quantity of magnetic energy in the Crab Nebula is readily estimated since the volume and the field strength are known. It is found to be 10^{47}—10^{48} erg.

Photographs of the Crab Nebula reveal a characteristic filamental structure. Other supernova remnants also show a similar structure. The synchrotron radiation is generated in the space between the filaments. The concentration of free electrons can be determined from their effect on the propagation of radio waves (the so-called Faraday rotation of the plane of polarization). It was found to be of the order of 1 electron per cm^3 in the Crab Nebula. Since the volume of the nebula is of the order of $5 \cdot 10^{55}\,cm^3$, there is a total of some 10^{56} electrons. There should also be a proportional number of atomic nuclei, mainly protons. Hence the total gas mass in the space between the filaments is of the order of 10^{32} g, i.e., $0.1\,M_\odot$. Note that for each 10^9 electrons there is on the average only one relativistic electron.

The radiation of the filaments is characterized by an emission spectrum corresponding to fairly high temperature, of about 10,000°K. The filaments occupy only a minor fraction of the nebular volume, but because of the fairly high gas density — of the order of 10^3 atoms per cm^3 — their total mass is also about $0.1\,M_\odot$. The factors responsible for gas emission in the filaments are not completely clear. This emission is possibly excited by synchrotron radiation in the ultraviolet, which is generated by relativistic electrons of extremely high energies.

Recent observations carried out beyond the atmosphere established that the Crab Nebula is also a powerful source of X-rays. This radiation component is possibly also of synchrotron origin. The frequency of the observed X-ray radiation is three-four orders of magnitude higher than the frequency of light. The energy of electrons generating synchrotron X-rays, according to equation (27), is therefore one order of magnitude higher than the energy of electrons which generate optical synchrotron radiation.

Study of the remnants of supernova envelopes thus led to highly important results. The explosion of a supernova generates a great swarm of relativistic particles and a large magnetic field. The total energy of the particles and the field even hundreds of years after the explosion is comparable in its order of magnitude to the overall energy emitted in the optical spectrum during the explosion. The energy released in the explosion also imparts high expansion velocities to the envelope, as in the case of novae, and it is this mechancial aspect of the explosion that we will consider now.

The envelope of a supernova expands through interstellar gas. The expanding envelope sweeps up the interstellar gas through which it moves, and the mass of the envelope is gradually built up. Since the velocities of the interstellar medium are very small compared to the expansion velocity of the envelope, the gas accretion apparently does not affect the initial momentum of the envelope. Since no external forces act on the envelope, its momentum remains constant while its mass increases, so that the velocity of the envelope steadily decreases.

Let m_0 be the mass of the envelope and v_0 its initial velocity. When the shell had receded to a distance r from the star, it will have swept up all the gas in a sphere of volume $4\pi r^3/3$. Hence the mass of the envelope will have increased by $4\pi r^3\rho/3$, where ρ is the density of the interstellar gas. The velocity v of the envelope at a distance r from the star can be obtained from the following equality, which is an expression of the law of momentum conservation:

$$\left(m_0 + \frac{4}{3}\pi r^3\rho\right) v = m_0 v_0. \tag{28}$$

The velocity v is clearly substantially less than v_0 only if the swept-up mass is comparable with the initial mass of the envelope. Let us apply equation (28) to the envelope of the star which exploded some 300 years ago in Cassiopeia. This envelope is now expanding at a velocity of nearly 7000 km/sec, i.e., it has not slowed down appreciably. And yet the mass of the swept-up gas is quite large. For an envelope radius $r \approx 7 \cdot 10^{18}$ cm and mean density of interstellar medium $\rho \approx 10^{-24}$ g it is of the order of the solar mass. The initial mass of the envelope was thus apparently several times the solar mass.

The Loop in Cygnus is another example of a supernova envelope of large mass. The present-day expansion velocity of the Loop is about 100 km/sec, and its radius is of the order of 10^{20} cm. The mass of the nebula determined from its volume and density (which as we know is calculated from the line spectrum) is equal to several hundred solar masses. This corresponds to the mass of the interstellar gas enclosed in the sphere of the above radius. Since the velocity v_0 of the envelope after the explosion was about 6000 km/sec (to judge from the typical envelope velocities of other type II supernovae), its mass m_0 according to equation (28) is found to be of the order of several solar masses. The explosion apparently occurred some 50,000 years ago.

The above examples and a number of other cases show that type II supernovae are capable of ejecting envelopes with masses of the order of $10 M_\odot$. The exploding stars themselves are therefore also highly massive. The kinetic energy of such an envelope is of the order of 10^{52} erg, which is thousands of times greater than the optical energy radiated by the star during the explosion. To realize what tremendous quantity of energy this is, remember that the Sun at its present power output will need several tens of billions of years to emit a comparable quantity of energy.

The kinetic energy of the nebulae produced by type I supernovae is substantially smaller, since the mass of these nebulae is one order of magnitude less than the mass of the envelopes of type II supernovae and their velocities are only about 1000 km/sec. This kinetic energy is of the order of 10^{49} erg, which is close to the estimate of the total radiant energy emitted by the star during the explosion.

What are the sources of the cosmic energies released in such tremendous quantities during supernova explosions? Clearly neither thermal nor radiant energy, even if instantaneously converted to mechanical energy, will ensure the observed velocities of the massive supernova envelopes. The total energy $M_* c_*^2$ of a star with a rest mass $M_* = 10-30\,M_\odot$ is of the order of $10^{55}-10^{56}$ erg. Nuclear fusion reactions synthesizing the heavy elements (up to iron) from hydrogen can release, as we shall in what follows, up to 1% of the rest mass in the form of energy. Therefore, conversion of a mere 10% of the total hydrogen reserve of a star with a $30\,M_\odot$ mass into heavy elements may in general supply the entire explosion energy of a supernova. As we know, however, nuclear reactions proceed very slowly under normal conditions and the instantaneous conversion of hydrogen into iron is unfeasible.

Another possible means of energy release is by a sudden contraction of the star. As the star radius changes from R_* to R_*', the potential energy drops from $-G\frac{M_*^2}{R_*}$ to $-G\frac{M_*^2}{R_*'}$. According to the virial theorem (see § 4), one half of the released energy, i.e., $\frac{1}{2}\left(G\frac{M_*^2}{R_*'} - G\frac{M_*^2}{R_*}\right)$ (again to orders of magnitude only) is converted to heat. If the star contracts sufficiently rapidly, the release of this thermal energy will become explosive. The various theories of supernova explosions therefore often start with the assumption of a collapsing star.

In accordance with one of the current theories of type II supernovae, the collapse is the result of the depletion of energy sources which sustain the star in its normal state. In § 4 we have described the proton-proton cycle which converts hydrogen into helium and supplies the energy requirements of the Sun. In the solar interior, where the density is of the order of $100\,\mathrm{g/cm^3}$ and the temperature is close to 15 million degrees, these reactions proceed at a fairly slow rate: more than ten billion years are required to convert a substantial fraction of hydrogen into helium.

The rate of thermonuclear reactions, however, rapidly increases with increasing temperature. In the interior of a massive star ($M_* = 10\,M_\odot$, say), the temperature is much higher than at the center of the Sun, reaching about 20 million degrees. Hydrogen therefore "burns" much more rapidly in a variety of reactions, which include both the proton-proton cycle and the so-called carbon cycle.

When most of the hydrogen has burnt out in the core of a massive star, the prevailing temperature (about 20 million degrees) is insufficient to trigger other nuclear reactions between heavy nuclei. The point is that the repulsion forces between heavy nuclei are greater than between protons, because the nuclear charge is greater than the proton charge. The particles must have higher energies in order to penetrate through the potential barrier, or in other words a higher temperature is needed.

When the bulk of hydrogen has burnt out, the star is left without a substitute energy source. It is gradually cooling by radiation and the gas density decreases. The equilibrium in the stellar interior is disturbed and the gravitation of the outermost layers, which is no longer balanced by the interior gas pressure, leads to a contraction of the star.

This contraction releases a certain quantity of potential energy which is converted to heat, so that the temperature and the density in the helium core of the star increase again. When the temperature rises to about

100 million degrees and the density to $10^5 g/cm^3$, He^4 nuclei (α particles) begin to be converted to heavier nuclei. The first stage is the formation of a beryllium nucleus from two α particles*:

$$He^4 + He^4 \rightleftarrows Be^8; \qquad (IV)$$

the beryllium nucleus fuses with another α particle to give a carbon nucleus:

$$Be^8 + He^4 \rightarrow C^{12}. \qquad (V)$$

Successive addition of α particles produces the heavier nuclei of oxygen, neon, and magnesium:

$$C^{12} + He^4 \rightarrow O^{16}, \qquad (VI)$$
$$O^{16} + He^4 \rightarrow Ne^{20}, \qquad (VII)$$
$$Ne^{20} + He^4 \rightarrow Mg^{24}. \qquad (VIII)$$

All the reactions (IV)—(VIII) are exothermal, i.e., they are accompanied by a release of energy.

Calculations show that by the time Mg^{24} starts forming, the helium resources of the star have been exhausted. Since the star again runs short of "nuclear fuel," the pressure drops and the core starts contracting. The contraction markedly raises the temperature. At temperatures of the order of three billion degrees (!), all the previously formed nuclei undergo thermonuclear fusion producing progressively heavier nuclei, mainly Fe^{56}. Nuclei with greater atomic numbers are also formed, but they are unstable, rapidly disintegrating. The core of the star is therefore now made up of iron.

The mass of the iron nucleus is 55.86 (on the scale of atomic weights), whereas the total mass of the 56 protons which go into the making of this nucleus is 56.46 atomic mass units. Hence, the chain of reactions starting with the formation of deuterium should release an amount of energy (including the energy carried off by the neutrino) equal to $\frac{56.46 - 55.86}{56.46} = 0.01$ of the total rest mass of the starting hydrogen. This is the maximum energy that can be released in nuclear reactions.

When the exothermal nuclear reactions have run to completion, the core of the star contracts again. When the central temperature is raised by contraction to 8 billion degrees, a special kind of nuclear reaction, associated with absorption of heat (an endothermal reaction), is triggered. This reaction involves the disintegration of Fe^{56} nuclei into α particles and neutrons. It proceeds at a very fast rate, and the core of the star cools rapidly. Since the pressure in the stellar interior falls, the star inevitably collapses toward the center. The collapse leads to a very marked increase in temperature. Vigorous nuclear reactions are now triggered in the parts of the star around the core where the nuclear fuel has not been exhausted yet, and an explosive release of energy occurs. The outermost skin of the star is stripped off by the explosion, and the event is observed as a type II supernova. Thus although the collapse of

* This nucleus is unstable, rapidly disintegrating into two α particles. A minute fraction of the nuclei, however, manages to fuse with α particles during their short lifetime, giving carbon (reaction V).

the star plays a significant role in setting the stage for the explosion, the eruption itself is a gigantic thermonuclear explosion.

Calculations based on this model were carried out for a stellar mass of $30\,M_\odot$ and an envelope mass of $10\,M_\odot$. Seeing that thermonuclear reactions in the explosive stage encompass a large part of the star, conversion of 1% of the total rest mass into energy can easily provide the observed energy of a type II supernova explosion.

We have considered in some detail only one of the theories of type II supernovae. Unfortunately the different theories still cannot be critically compared and evaluated because of lack of relevant information. For type I supernovae, whose pre-explosion masses are of the order of one solar mass, the theory also assumes a thermonuclear explosion with the participation of heavy nuclei, but here the ground is even less certain.

The results of observations of supernova explosions led to considerable advances in the understanding of cosmic rays, whose origin is one of the most attractive problems of modern physics. Although the opinions vary as to the exact mechanism which accelerates the particles to velocities close to the velocity of light, the role of supernova explosions as a source of cosmic rays has been firmly established.

High-energy electrons are apparently not the only kind of relativistic electrons produced by a supernova explosion. So far, we are unable to detect relativistic protons and heavier nuclei in supernova remnants, because their radiation in the weak magnetic fields of the nebulae is most inconspicuous. It seems, however, that the energy of all the relativistic particles in supernova envelopes is one order of magnitude higher than the energy of the electrons. The total energy of cosmic rays in a nebula is generally assumed comparable with the magnetic energy. If the field energy were much less than the cosmic ray energy, the particles would simply slip out from the nebula. They are confined inside the nebula only because the magnetic field curves their trajectories.

The total energy of relativistic particles in the Crab Nebula is currently estimated as $5 \cdot 10^{48}$ erg, and in some other supernova remnants it is one order of magnitude higher. This energy is gradually dissipated as radiation. Moreover, the nebula expands, its volume increases, and the magnetic energy density diminishes. The field strength thus decreases, and the combined effect of these two factors should gradually reduce the radiation flux from relativistic particles.

The electrons responsible for the X-ray and the optical radiation of the nebula lose their energy particularly rapidly. Since relativistic electrons are still observed in the Crab Nebula, we conclude that some source continues generating relativistic electrons for a long time after the explosion. In the 900 years which elapsed after the explosion, the original "optical" and "X-ray" electrons will have lost all their energy and can no longer emit at these frequencies. At this stage we do not know what is the exact source of relativistic electrons, whether the star or the nebula, since the mechansim of their formation is not known.

Let us compare the concentration of cosmic rays in the interstellar space with the concentration produced by supernova explosions in the Galaxy. Measurements of the number of relativistic particles reaching the Earth's atmosphere (these measurements were carried out from rockets) show that the cosmic ray energy is of the order of 10^{-12} erg per cm^3 of space. When in space, the particles lose their energy in

10^{16} sec in collisions with the atoms of the interstellar gas. The total volume of the Galaxy ($10^{68}-10^{69}$ cm^3) contains $10^{56}-10^{57}$ erg of energy in the form of cosmic rays. Thus, each second some $10^{40}-10^{41}$ ergs are lost.

If we remember that a supernova explodes once every 300 years, forming cosmic rays with a total energy of the order of 10^{50} erg, the addition to the total energy of cosmic rays in the Galaxy is on the average 10^{40} erg per second, and this figure is close to the rate of energy losses in collisions. Supernovae are thus capable of continuously replenishing the cosmic ray energy in the Galaxy. Whether or not they are the main source of cosmic rays is not clear at this stage, since our knowledge of the diffusion processes of cosmic rays from the Galaxy into the intergalactic space is still fragmentary. Anyhow, a large proportion of particles reaching the Earth's atmosphere, where they fission the nuclei of air atoms, originated some time in the past in the interior of a supernova and traveled through a maze of interstellar magnetic fields until they encountered the Earth.

§ 8. EXPLOSIONS IN GALACTIC NUCLEI

For a long time supernova explosions were regarded as the topmost rung on the scale of cosmic catastrophes. Recently, however, much more powerful explosions were discovered, releasing energy equivalent to a few million solar masses. Such tremendous explosions clearly cannot occur in individual stars. They take place in the central regions (nuclei) of galaxies, whose masses are equal to billions of solar masses. In this section we will discuss the explosions in galactic nuclei, but first we will examine in more detail than so far the galactic universe itself.

Numerous galaxies are spirals, not unlike our Galaxy. On the other hand, there are also star systems of regular elliptical form, without any prominent spiral arms. These are the elliptics. The third category comprises the so-called irregular galaxies, which are devoid of any specific shape: these are amorphous nebulous objects. The two nearest neighbors of our Galaxy — the Large and the Small Magellanic Clouds — are irregular galaxies. These three principal types are further subdivided into a range of subtypes.

Individual bright stars, globular clusters, and bright gas nebulae are distinguishable in the nearest galaxies. For analogous objects in our Galaxy, we know the relation between the apparent birghtness (stellar magnitude) and the distance from the Sun, so that by extrapolation we can find the distances of various galaxies from observations of the apparent stellar magnitudes of the individual visible objects. In this way, a certain average integrated brightness was assigned to galaxies of each subtype.*

The apparent brightness of a galaxy also depends on its distance from the Sun: it is inversely proportional to the square of the distance. Therefore, knowing the type of the galaxy, we can find its average distance from the integrated brightness even if no individual stars are observable.

* The integrated brightness of a galaxy is the total brightness of all the constituent stars.

The spectra of very weak and extremely distant galaxies were obtained during the 1920s with the powerful 200-in. telescope, and these studies revealed a most remarkable feature. The lines in the spectra of distant galaxies are shifted toward longer waves (to the red), and this red shift is proportional to the distance of the galaxy from the Sun. The red shift law is expressed by the equality

$$\frac{\Delta\lambda}{\lambda} = \frac{H}{c} r, \tag{29}$$

where $\Delta\lambda$ is the wavelength shift, λ is the unshifted wavelength, r is the distance of the galaxy, and H is the so-called Hubble constant. Modern observations give for H a numerical value close to 30 km/sec per 1 million light years. This means that in the spectrum of a galaxy distant 300 million light years a line of wavelength λ is shifted by $\Delta\lambda = \frac{3 \cdot 10^{\cdot}}{3 \cdot 10^{\cdot}} \cdot 3 \cdot 10^2 \lambda = 3 \cdot 10^{-2} \lambda$.

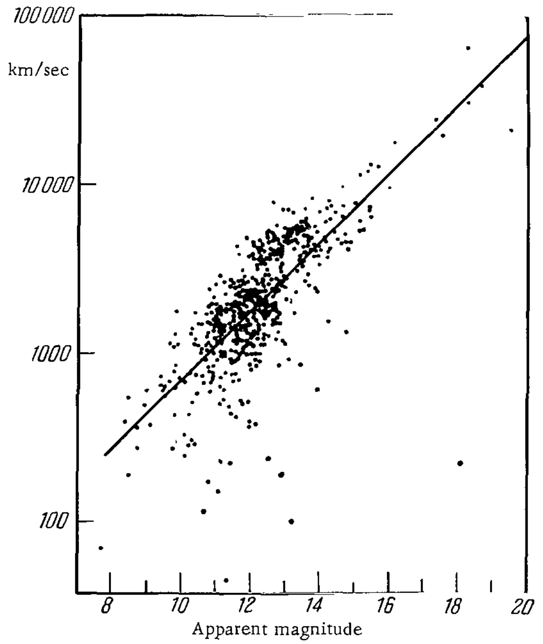

FIGURE 39. Relation between the apparent stellar magnitude and the red shift. (The higher the apparent magnitude, the farther is the galaxy.)

The red shift in the spectra of distant galaxies is associated with the Doppler effect: it is the outcome of a special kind of motion of the galaxies, which all seem to recede from the observer. This, however, is not to be interpreted as an indication of the central position of our Galaxy in the Metagalaxy. The situation is much simpler: all the galaxies recede from one another, and an observer situated in any of these galaxies will see all the other galaxies (our Galaxy included) flying away from him.* The exact nature of this phenomenon has not been established, and we speak to the "red shift" only as a convenient device for measuring the distances of the farthest objects in the Metagalaxy.

On the average, the largest galaxies are the spirals. They also contain a greater stellar population than galaxies of other types. Among the spirals are real giants, such as our Galaxy. The smallest dwarf galaxies contain only tens of millions of stars. These are usually ellipticals.

The galactic masses are calculated in the same way as the mass of the Galaxy, namely from the rotation velocity (see §3). The mass of our Galaxy from rotation data was found to be close to $2 \cdot 10^{11} M_{\odot}$. This is the characteristic mass of the giant galaxies. In most cases, however, the galactic masses are about $10^9 - 10^{10} M_{\odot}$.

The very bright central part of a galaxy with a diameter of a few tens of light years is called the galactic nucleus. On photographs, nuclei are distinguishable only in the nearest galaxies, e.g., the Andromeda Nebula (M 31, the thirty-first galaxy in Messier's catalogue). But even for the nearest galaxies, the fine structure of the nucleus cannot be resolved. It has been generally assumed that the nuclei of spiral galaxies are dense condensations of stars, but recent observations seem to point to a more complex structure.

The nucleus of our Galaxy is naturally the best place to begin the research of galactic nuclei. Unfortunately, it is obscured by dense absorbing gas-dust clouds, so that even the part of the Galaxy around the nucleus are not visible. The region of the galactic nucleus was studied by radio astronomical methods, and some results of these observations are presented in what follows.

Most of the stars of a spiral galaxy are contained in a central disk. The spiral arms are populated with hot giants of the spectral types O and B and they also carry various bright gaseous nebulae, which are regions of interstellar gas illuminated by these hot stars. Closer to the center, the proportion of cold giants is higher. The space between the arms is by no means empty, but the density of stars and gas there is much less than in the spiral arms.

We will not go into the problem of the optical radiation of the galaxies, since it is determined by the thermal emission of the stars and the interstellar medium, discussed in considerable detail in the previous sections. Besides the optical light, however, the galaxies radiate at radio frequencies. Our Galaxy, for instance, is also a source of radio waves. It is only the centimeter and decimeter waves that are emitted mainly by hot gas, whereas the longer radio waves are mostly synchrotron radiation. It is emitted by relativistic electrons moving in interstellar magnetic fields.

An observer situated outside our Galaxy will see it as a fairly weak radio source: it emits hundreds of thousands less in the radio spectrum than in the optical spectrum. However, the radio flux of some star systems is thousands and tens of thousands of times more powerful than that of our Galaxy and other normal star systems. These strong radio sources are known as radio galaxies.

* A good model of this is the motion of points on the surface of a balloon when it is being blown up. None of these points is in a preferential position, but all the points seem to recede from any chosen point.

In a number of cases radio galaxies could be identified with optically visible systems. There are radio sources, however, which are simply invisible in the optical light, and we then refer to a discrete radio source. The angular dimensions of the optical object identified with a radio galaxy are generally much less than the size of the radio source. This indicates that the bulk of the galactic mass emitting both in the optical and in the radio spectrum is surrounded by a very extended envelope which gives no visible light. These radio envelopes are observed for some normal galaxies as well, although their emission is extremely weak.

If we assume that the radio waves from radio galaxies are of thermal origin, the observed radio brightness corresponds to gas temperatures of billions of degrees. At these high temperatures the optical radiation should be incomparably stronger than the radio emission. And yet the radiation power of radio galaxies in the radio spectrum is comparable with the optical power. This leads to the conclusion that the radio waves from these galaxies are mainly of nonthermal origin. Various data seem to indicate that these radio waves, like the long radio waves of our Galaxy, are generated by the synchrotron mechanism. One of the principal facts supporting this point of view is the polarization of the radiation from radio galaxies, both in the optical and in the radio spectrum.

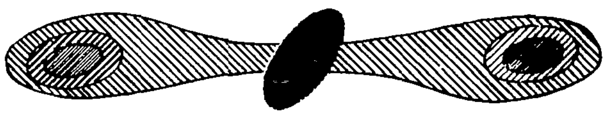

FIGURE 40. Schematic structure of the radio source Cygnus A. The optical object is shown at the center: this is a galaxy with a double nucleus. The radio source is represented by the cross hatched area.

The information that we obtain from observations of radio galaxies proved most useful for the study of explosions in galactic nuclei. First let us describe the radio galaxy in the constellation of Cygnus, known as Cygnus A. In 1954 an optical photograph of this extremely powerful extragalactic radio source was obtained. The red shift of the lines in the spectrum of the optical object gave a distance of some 500 million years for Cygnus A from the Sun. Once the distance has been found, the observed radio flux from this galaxy could be used to calculate the total radio energy emitted by the source. The result gave an estimate of the order of 10^{45} erg/sec, which is much greater than the combined emission of our Galaxy in both the optical and the radio spectrum. The visible image of Cygnus A is relatively weak, and the radiation in the optical spectrum is only a small fraction of this figure.

The most remarkable feature of Cygnus A, which immediately attracted the astronomers' attention, is its binary structure. It consists of two extended radio sources whose centers are separated by some 500 light years, and between them there is an optically visible object of about 1/10 the size of the radio source. This optical region in its turn is made up of two parts. The radio source Cygnus A thus can be regarded as a galaxy with a double nucleus. Two giant plasma formations move in opposite directions from the nucleus with velocities of several thousand kilometers per second.

The galaxy in Cygnus A contains tremendous gas clouds which move at random with high velocities. This conclusion is based on observations of the optical spectrum of the galaxy, which shows a multitude of emission lines characteristic of gaseous nebulae. The width of these lines indicates that they originate in gas moving at random with velocities of up to 500 km/sec.

For some time after the discovery of the binary structure of the source Cygnus A, the peculiar formation was generally regarded as the result of two giant galaxies in collision. This point of view has been abandoned by now, mainly because it does not explain the tremendous power of the source. When galaxies collide, only a minor fraction of the total reserve of kinetic energy is converted to radio waves. Moreover, the odds against a collision between two giant galaxies, like those observed in Cygnus A (the colliding nuclei are very small indeed) is fantastic. If this extremely unlikely head-on collision did occur, however, we can expect much more numerous glancing collisions between dwarf galaxies, but none are observed.

There is, however, an entirely different approach to the interpretation of Cygnus A and other radio galaxies. According to this school of thought, the nucleus of the galaxy associated with Cygnus A exploded some time ago, and thus fissioned into two components. Two gas jets were ejected at the same time into opposite directions, and it is these formations that are now observed as the radio sources.

The age of the radio galaxy Cygnus A, i.e., the time elapsing after the explosion, can be estimated in a variety of ways. It is at least 10^3 years, but probably it is much higher, reaching 10^6-10^7 years. The power of the radio emission from this source is currently 10^{45} erg/sec or higher, and it could hardly have been less immediately after the explosion. The energy released by the explosion and the subsequent processes is therefore at least $10^{56}-10^{58}$ erg. Since we observe only the radiation in isolated regions of the spectrum and since it is reasonable to assume that initially the radiation was even more powerful than it is now, the explosion energy is raised to $10^{59}-10^{60}$ erg.

The structure of some other powerful extragalactic radio sources, e.g., Centaurus A, Fornax A, is highly similar to that observed in Cygnus A. These are double radio galaxies, and the centers of radio emission are situated symmetrically about the optical galaxy, at a considerable distance from it. In all cases, an explosion in the galactic nucleus caused ejection of galactic matter in two opposite directions with roughly equal momentum.

Effects associated with explosive processes which encompass a large part of the star system are also observed in other galaxies, whose binary structure is not immediately obvious. A highly remarkable example in this connection is the giant elliptical galaxy M 87, which is distant 50 million years from the Sun. This system is observed in the constellation of Virgo, and it actually coincides with the very powerful radio source Virgo A.

The photographs of M 87 (Figure 41) clearly show a luminous formation, a jet or a stream extending from the central part of the galaxy. This jet is made up of several condensations, whose optical radiation is highly polarized. The length of the jet is several thousand light years. Its color is blue, and the spectrum contains no emission or absorption lines. The separation of the principal condensations in the jet from the center of the galaxy is at least several tens of light years.

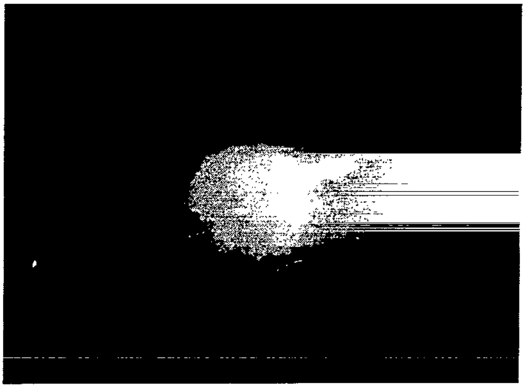

FIGURE 41. The galaxy M 87 (radio source Virgo A). Note the ejection from the galactic nucleus on the right.

The connection between the jet and the nucleus of M 87 is obvious, and there can be no doubt that the jet was indeed formed as a result of an explosive process in the nucleus. The ejection of gas from the nucleus of M 87 goes on to this day, as is evident from its spectrum. The spectrum of the part of the galaxy near the nucleus shows emission lines shifted in the direction of short waves. This blue shift is apparently associated with the motion of luminous gas masses. Shift measurements give velocities of the order of several thousand kilometers per second.

The radio waves originate both in the galactic nucleus and in the surrounding region which extends over a distance of the order of a hundred thousand light years. Strong radio emission, especially at short (decimeter) waves, is also observed from the jet. Since both the optical and the radio emission of the jet are highly polarized, it is definitely associated with synchrotron radiation. As in the Crab Nebula, the optical radiation is an extension of the radio spectrum in the direction of short waves.

An estimate of the magnetic field strength in the jet gives values of the order of 10^{-4} oersted. The high-energy electrons responsible for the optical radiation of the jet will lose their energy in these fields in approximately a thousand years. And yet the jet has existed at least for several tens of thousands of years, if the ejection velocity is assumed to be close to the velocity of light. In all probability the explosion in the nucleus took place a few million years ago. The relativistic electrons producing the optical emission of the jet therefore were not ejected from the nucleus, and acquired their relativistic energies in the jet itself. We thus conclude that the explosion in the nucleus of M 87 led to the ejection of a peculiar formation, which to this day remains a source of relativistic electrons.

The radiation power of the jet is of the order of 10^{43} erg/sec. During the several million years that have elapsed since the explosion, it apparently radiated a total of $10^{56}-10^{57}$ erg. We have thus obtained the same lower-bound estimate for the energy of an explosion in the galactic nucleus as the previous estimate for Cygnus A and other sources. The quantity of energy released in these explosions in tens of millions of times higher than the energy of a supernova explosion.

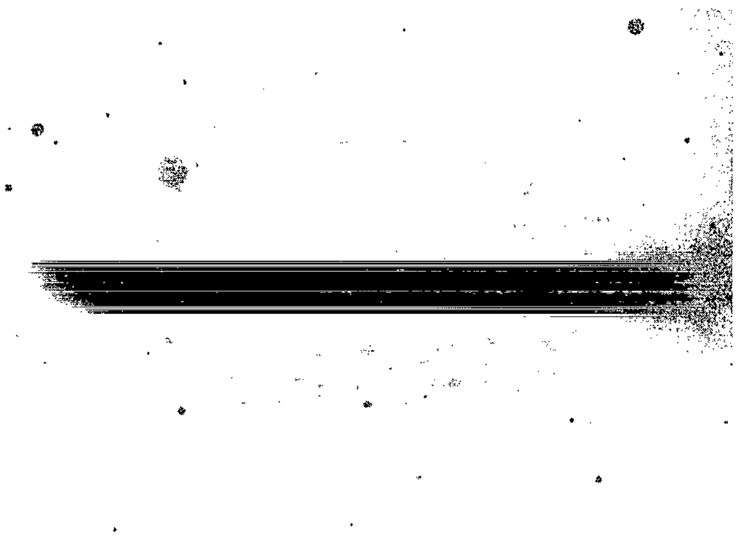

FIGURE 42. The galaxy M 82 (photographed in continuous light).

Observations of the relatively near irregular galaxy M 82 revealed an extremely interesting pattern of gas motions associated with a recent explosion in the galactic nucleus. In this galaxy, despite its irregular shape, we can distinguish between two preferred directions — one along the maximum longitudinal dimension of the galaxy and the other perpendicular to it (Figure 42). We will refer to these directions as the major and the minor axis. A system of filaments is observed in the direction of the minor axis of M 82. They emit mainly at line frequencies, and not in the continuous spectrum, and the strongest emission line is the red hydrogen line H_{α}. Photographs of the nebula taken with an optical filter which transmits in tne light of H_{α} and in a narrow frequency interval around it give an excellent picture of the filaments. Comparison of Figures 42 and 43 reveals further differences between the regions emitting mainly in the line spectrum and the regions of continuous emission. The filaments extend over 10,000—12,000 light years from the center of the galaxy.

The line shifts observed in the spectra of the filaments show that the gas in the filaments moves with velocities of about 1000 km/sec from the center. Three million years are needed to cover a distance of 10,000 light years while moving with this velocity. The explosion in the nucleus of M 82 thus took place a few million years ago.

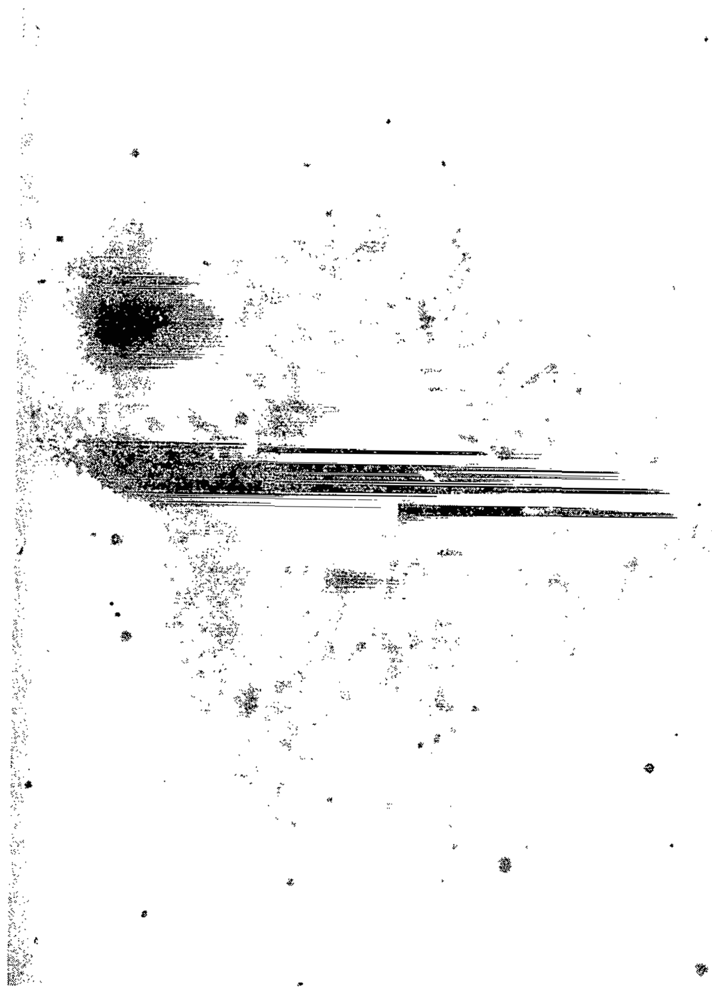

FIGURE 43. The galaxy M 82 (photographed in H_α line). Note the filamental structure at the center.

With regard to its filamental structure, M 82 is not unlike the Crab Nebula. This likeness is further emphasized by the strong polarization of the light from the filaments of M 82. Finally, as in the case of the Crab Nebula, the filamental region of M 82 is also a radio source (though not very powerful).

In the light of these findings we are led to the conclusion that the continuous radiation of M 82 is of synchrotron origin. The characteristic arch-like shape of the filaments (see Figure 43) is apparently attributable to the effect of magnetic fields: the plasma moves along the field lines. The polarization observations helped to establish the direction of the magnetic field lines, and it was found to coincide on the whole with the prevailing direction of the filaments.

The line emission of the filaments of M 82 is explained as that of the Crab Nebula. The extremely energetic relativistic electrons in the filaments apparently radiate photons corresponding to the ultraviolet part of the spectrum. These photons excite the gas atoms and are thus indirectly degraded to emission at line frequencies. M 82 also emits X rays, which suggests that the galaxy probably contains electrons of even higher energies.

Although the structure of the central parts created by the explosion in the nucleus of M 82 is superficially similar to nebulae produced by supernova explosions, the two effects are entirely different in terms of scale. The energy E_0 radiated by the galaxy at the H_α line frequency, which reaches the earth-bound observer is of the order of $2 \cdot 10^{-11}$ erg/cm$^2 \cdot$ sec. Since the distance r to the galaxy is about 25 million light years, the radiant power emitted in the H_α line is $4\pi r^2 E_0 \approx 10^{41}$ erg/sec. Note that the radiation of M 82 remains extremely powerful millions of years after the explosion, whereas the Crab Nebula a mere 900 years after the explosions emits only about 10^{34} erg/sec.

Let us calculate the kinetic energy of the gas which moves away from the nucleus of M 82. The mass of this gas is calculated from its volume and density. The volume measured from photographs of the galaxy is of the order of 10^{63} cm^3. The concentration of hydrogen atoms in the emitting gas was estimated from the observed H_α radiation flux; it is of the order of 10 atoms in 1 cm^3. Hence, the total number of atoms in the volume of the galaxy is approximately 10^{64}, and the total mass of the hydrogen gas is about $2 \cdot 10^{40}$ g. We have mentioned before that the filaments move with a velocity of some 10^8 cm/sec, and the kinetic energy is thus of the order of 10^{56} erg.

The total quantity of energy released by the explosion in the nucleus of M 82 was partly converted into the kinetic energy of the gas and partly into the energy of cosmic rays and the magnetic field, which is currently estimated as $10^{55}-10^{56}$ erg. Moreover, the energy radiated by the galaxy during the period which elapsed after the explosion is at least 10^{55} erg, and possibly 10^{56} erg. We thus obtain for the explosion energy of M 82 a figure of the order of $10^{56}-10^{57}$ erg, which virtually coincides with the explosion energy in the nuclei of other galaxies.

The explosion in a galactic nucleus thus produces vigorous motion of the gas in the central part of the galaxies. Of considerable interest are the so-called Seyfert galaxies (named after their discoverer), where the nucleus is characterized by extraordinary activity. These active nuclei are significantly brighter than the rest of the galaxy. Moreover, the spectra of the nuclei of Seyfert galaxies contain emission line of ionized atoms of various elements. The lines are very wide and reveal complex structure. They are made up of numerous "spikes." This structure suggests that the lines form in giant assemblies of gas clouds moving at random. Since the emitting gas masses move in different directions, their line-of-sight velocities also differ. The emission lines from different clouds therefore show different wavelength shift, and the overall result is a highly broadened and "spiky" emission line. Line broadening points to velocities of from 500 to 3000 km/sec in the nucleus.

One of the best known Seyfert galaxies (some ten have been discovered so far) is the spiral NGC 1068 (NGC is an abbreviation for the New General Catalogue of galaxies, and 1068 is the running number in the catalogue). The distance of this galaxy is some 40 million light years. The nucleus does not show on positive prints: it is only distinguishable on the original negative as a bright star at the center of the galaxy; the nucleus is only 100 light years across or even smaller. A bright region is visible at the center of the photograph: these are the luminous gases surrounding the nucleus. They also give a spectrum of emission lines. However, the lines in the spectrum of the nucleus proper are very wide, whereas the surrounding gas envelope gives much narrower lines. The fast moving gases are therefore concentrated in the nucleus proper.

The properties of the gas in the nuclei of Seyfert galaxies — its chemical composition, density, and temperature — have been repeatedly determined from their line spectra. The nuclear gas in NGC 1068 was found to consist mainly of hydrogen with an average concentration of the order of 10^3 atoms per cm^3 and a temperature of about 20,000°. The gas clouds are apparently distributed in a most nonuniform manner through the nucleus of the galaxy, and their total volume is $10^{60}-10^{61}\,cm^3$. The mass of the gas in the nucleus is on the average $10^6 M_\odot$ (possibly reaching $10^7 M_\odot$ in other galaxies), and the kinetic energy is of the order of $10^{55}-10^{56}$ erg. We have previously obtained a figure of the same order of magnitude for the nucleus of M 82. The vigorous motions in the nuclei of Seyfert galaxies are probably also created by explosive processes. All the alternative explanations of this activity (e.g., thermonuclear reactions) meet with insurmountable difficulties.

The gas clouds move at random and thus collide repeatedly. Because of their tremendous velocities, the collisions raise the temperature of the gas as a certain proportion of the kinetic energy of the clouds is converted to heat. The observed line spectrum of Seyfert galaxies is indeed the emission spectrum of hot gas. At line frequencies the nucleus emits some $10^{42}-10^{43}$ erg/sec. If the entire kinetic energy of the clouds were converted into radiation, the emission would continue only for 10^{13} sec, i.e., for a few hundred thousand years. In practice, only a minor fraction of the kinetic energy is converted to visible radiation, and therefore the kinetic energy cannot sustain the luminosity of the nucleus even for this short period. On the other hand, we know that the explosion in the nucleus of any Seyfert galaxy occurred at least a few million years ago. Indeed, gas moving with velocities of the order of 1000 km/sec from the explosion region takes millions of years to cover the distance equal to the radius of the luminous volume — $10^{21}-10^{22}$ cm. It therefore seems that either energy is constantly "pumped" into the gas from other sources or else its original kinetic energy was much higher than it is now. If the latter were true, however, we would obtain explosion energies much in excess of the above estimate of $10^{55}-10^{56}$ erg.

There are stars, as well as gas clouds, near the center of Seyfert galaxies. These stars produce the characteristic stellar absorption lines in the spectrum. The lines form in the spectra of individual stars, and they show in the resultant spectrum of the galaxy only because the radiation of the stars of any given type is deficient at the line frequencies. The observed continuous emission of the nuclei of Seyfert galaxies is also produced by stars, and it is a factor of 5—10 stronger than the total line emission. However, since line emission is distributed between relatively few and comparatively narrow spectral intervals, the radiation flux in each interval is sufficiently high for the line to be clearly visible against the continuous background.

The nuclei of some Seyfert galaxies, and in particular that of NGC 1068, also emit a fairly strong radio flux of synchrotron origin. The activity of these nuclei apparently involves generation of relativistic particles, as well as turbulent gas motion. The total energy of relativistic particles in the nucleus of NGC 1068 is of the same order of magnitude as the kinetic energy of the gas, or only slightly less.

To summarize, we have considered different types of star systems — galaxies — with active nuclei. This activity mainly results in strong radio-frequency emission from the nucleus or in ejection of gas from the nucleus. There are also galaxies with turbulent motion of gas in and near the nucleus. In all cases, the activity can be attributed to an explosion which took place in the galactic nucleus hundreds of thousands or millions of years ago. The explosion released a tremendous amount of energy, no less than $10^{56}—10^{57}$ erg in various forms.

The above examples clearly do not cover the entire range of activities in galactic nuclei. It is also clear that further development of the astronomy of extragalactic objects will lead to the discovery of new forms of nuclear activity in galaxies. It should be remembered, however, that explosions in galactic nuclei is not a frequently recurring process, and the effect of each explosion naturally persists for a time which is very short compared to the entire lifetime of a galaxy. For the rest of the time, the activity of galactic nuclei may remain quite low, and it is therefore observable only for the nearest galaxies.

The nucleus of our Galaxy also shows signs of activity. We have mentioned in passing that the central parts of the Galaxy are inaccessible to optical observations because of thick obscuring dust. Some information on the structure of the nucleus has been obtained by radio observations, since radio waves propagate unimpeded through interstellar dust. Two very strong radio sources are located at the center of the Galaxy, each with a diameter of some ten light years. The spectrum of these radio sources identifies them as sources of synchrotron radiation. The nucleus of our Galaxy is thus fundamentally similar to the nuclei of other galaxies.

Gas is ejected from the central regions of the Galaxy with velocities of about 100 km/sec. In one year, the nucleus ejects about one solar mass. It thus seems that the gas concentration in the nucleus is very high compared to the other parts of the Galaxy: despite its small dimensions, the nucleus must have the mass of a million Suns at least.

Examination of the probable nature of galactic nuclei and their contribution to the evolution of galaxies is postponed until § 10. Here we will only briefly consider to what extent the known energy sources are capable of releasing some $10^{56}—10^{57}$ erg in a relatively short time.

The collision hypothesis is definitely inapplicable to the energy release in radio galaxies and other galaxies with exploding nuclei, since the activity is very often observed in the nuclei of solitary galaxies also. The reason for explosions is to be sought inside the nuclei themselves.

Conversion of potential energy to other forms of energy in a contracting star system does not answer the case, either, since the tremendous dimensions of the galaxies preclude all possibility of catastrophic collapse. Moreover, it has been firmly established that the explosions are localized inside the very small volumes occupied by the galactic nuclei.

Tremendous difficulties arise when thermonuclear reactions are involved in order to account for the explosive processes in the galactic nuclei. If we adopt the thermonuclear hypothesis, we have to assume that the small nucleus contains a tremendous number of stars which erupt as supernovae in an unusually rapid succession — on the average once every year. The reasons for these frequent supernova eruptions are not at all clear, and there is no observational evidence that the concentration of stars in galactic nuclei is unusually high. Moreover, the supernova hypothesis does not shed light on the asymmetric, one-sided ejection of matter from galactic nuclei, as in M 87.

The discovery of explosions in galactic nuclei thus necessitated a complete revision in our approach to the conversion of mass into energy. To review the present state of this problem, we now pass to another class of extremely interesting objects, known as quasars (which is an abbreviation for quasistellar radio sources). On the energy scale, they are hundreds and thousands of times more powerful than even the explosions in galactic nuclei. Therefore, although it is not entirely clear that we are dealing with explosives in the case of quasars, their study is of the utmost significance for understanding the nature of cosmic explosions.

§ 9. QUASARS

The discovery and the study of quasars was made possible only by the steady development of radio astronomy, which took place since the late 1940s. Large radio telescopes pinpointed numerous discrete radio sources in the sky. Most of the sources lying outside our Galaxy are independent radio galaxies, and some of these we described in the previous section. Radio galaxies (even those which have not been identified with optical objects so far) are extended formations with easily measurable diameters.

In 1963, five powerful discrete radio sources were discovered, which could not be classified as radio galaxies: their angular dimensions proved very small compared to the average size of radio galaxies. Since they are not unlike stars in this respect, they were called quasistellar radio sources, or quasars in short. Some fifty quasars have been listed in 1966. They are designated by their running number in the catalogue of discrete radio sources, e.g., 3C 273 is to be read as source number 273 in the Third Cambridge Catalogue.

The 200-in. telescope enabled the astronomers to identify some of the quasars with optical objects and to photograph their optical spectrum. Quasars appear as faint stars surrounded by nebulosities which are hardly distinguishable on photographs. The spectra of quasars show wide emission lines. These are the lines of hydrogen, oxygen, magnesium, and other common nebular elements, but they are all markedly shifted toward the red. The ratio of the line shift $\Delta\lambda_0$ to the wavelength λ_0 is found to be constant for all the lines in the spectrum of a given quasar. For the quasar 3C 273 this ratio is 0.16, and for 3C 48 is 0.37, and even higher for other quasars. Because of this shift, the spectral lines of quasars look highly unusual. For example, the H_β hydrogen line, which is normally located in the green part of the spectrum, lies in the red region in the spectrum of 3C 48, and so on.

There can be two alternative interpretations of the position of quasars in space and two different explanations of the origin of the red shift in their spectra. One possibility is that the quasars are extragalactic objects located at tremendous distances from the Sun. In this case, the shift in their spectra is apparently of the same origin as the red shift of all the distant galaxies. If, on the other hand, the quasars lie in the Galaxy, the line shift in their spectra cannot be attributed to the Doppler effect: if the quasars indeed had velocities of the order of a hundred thousand kilometers per second, their motion would be clearly visible against the background of other stars, whereas observations do not reveal anything of the sort. If the quasars are galactic objects, the red shift in their spectra should thus be assigned to some other effect. Gravitation is another possible cause of the red shift, and we will now describe how it explains the quasar spectra.

A photon escaping from the surface of a star overcomes its gravitational pull, i.e., it does work against the force of gravity. If the photon energy is $h\nu_0$, its equivalent mass according to equation (18) is $\frac{h\nu_0}{c^2}$. The force F acting on this mass on the surface of a star of radius R_* is

$$F = G\frac{M_*}{R_*^2}\frac{h\nu_0}{c^2},$$

where M_* is the mass of the star. The force on the photon varies as the photon moves away from the star, and exact calculation requires application of calculus. However, reasoning along the same lines as in §4, we can easily obtain a working estimate.* The work done to remove the photon from the surface of the star is

$$A = \frac{GM_*}{R_*}\frac{h\nu_0}{c^2}.$$

Since the escaping photon does work, its energy diminishes by the amount A. The energy $h\nu$ of a photon reaching the observer is related to its initial energy $h\nu_0$ by the equation $h\nu = h\nu_0 - \frac{GM_*}{R_*}\frac{h\nu_0}{c^2}$. The decrease in the photon frequency $\Delta\nu_0 = \nu - \nu_0$ is thus given by

$$\frac{\Delta\nu_0}{\nu_0} = -\frac{GM_*}{R_*c^2}. \qquad (30)$$

* We assume that the force remains constant over a distance R_* and then vanishes. By using a larger force over the entire range, we compensate to a certain extent for the neglect of the work done over distances greater than R_*.

The corresponding change in wavelength $\Delta\lambda_0 = \lambda - \lambda_0$ (seeing that $\lambda_0 = \frac{c}{\nu_0}$) is thus

$$\frac{\Delta\lambda_0}{\lambda_0} = \frac{\frac{GM_*}{R_* c^2}}{1 - \frac{GM_*}{R_* c^2}}. \tag{31}$$

For photons escaping from the surface of a normal star, $\frac{\Delta\lambda_0}{\lambda_0}$ is very small. In the case of the Sun, whose mass is $2 \cdot 10^{33}$ g and radius $7 \cdot 10^{10}$ cm, we have $\frac{\Delta\lambda_0}{\lambda_0} = 2 \cdot 10^{-6}$. This change in line wavelengths is hardly noticeable. For $\frac{\Delta\lambda_0}{\lambda_0}$ to be of the order of 0.1, the entire solar mass should be condensed into a sphere some 10 km in radius. The brightness of the star is proportional to the square of its radius. Suppose a star of radius 10 km and temperature of the order of 10,000° (these are the temperatures inferred from the spectra of quasars) has the same magnitude as that of the quasar 3C 273. This star should be much closer to the Sun than all the other known stars, and will appear to move across the sky owing to the Sun's proper motion. In fact, however, the quasars maintain fixed positions, so that their distances are at least 60,000 light years.

The conclusion that the quasars are distant objects is also confirmed by radio observations. The radio waves from quasars (wavelength $\lambda = 21$ cm) experience the same absorption in the interstellar medium as the radio waves from radio galaxies. The radio waves from quasars thus have to cross the entire galaxy on their way to the Sun, which indicates that they are no nearer than the rim of the Galaxy and possibly even farther off. To make these distances consistent with the apparent brightness of quasars and the amount of the red shift from equation (31), we have to assume a radius of the order of 10^{15} cm and a mass of $10^9 M_\odot$. Theoretical analysis suggests, however, that stars with such radii and masses simply cannot exist.

It thus clearly follows that gravitation cannot account for the red shift in the spectra of quasars.* We are thus left with the only possibility, namely that quasars lie outside the Galaxy, and the red shift is associated with the overall expanding motion of the Universe. We have discussed in some detail the facts behind this conclusion, since it is of fundamental importance for understanding the nature of quasars. The assumption of the extremely high luminosity of quasars (which was one of the main reasons for the increased interest that scientists started to show in quasars) is based mainly on the interpretation of the red shift as a result of the Doppler effect.

The relation of the line shift $\Delta\lambda_0$ to the velocity v is expressed by equation (5) only if v is small compared to the velocity of light ($v \leqslant 0.1\,c$ say). At higher velocities, the expression of the Doppler effect should be

* Note that for stable superdense stars this formula is applicable only up to $\frac{\Delta\lambda_0}{\lambda_0} \lesssim 0.3$, and in general $\Delta\lambda_0$ can never exceed $0.6\,\lambda_0$. Since there are quasars showing $\frac{\Delta\lambda_0}{\lambda_0} > 1$, the red shift cannot be due to the gravitational pull of a stable star.

taken from the theory of relativity:

$$\frac{\Delta\lambda_0}{\lambda_0} = \frac{2\frac{v}{c}}{\sqrt{1-\frac{v^2}{c^2}}+1-\frac{v}{c}}. \tag{32}$$

Clearly, when $\frac{v}{c}$ is small compared to 1, equation (32) gives virtually the same result as the simpler equation (5). If, however, $\frac{v}{c}$ is sufficiently close to 1, $\frac{\Delta\lambda_0}{\lambda_0}$ may become inordinately large.

Taking the velocity v to be proportional to the distance of the quasar (with a proportionality coefficient equal to Hubble's constant), we find that the quasar 3C 273 with $\frac{\Delta\lambda_0}{\lambda_0} = 0.16$ is distant about 1.8 billion light years. The distances of other known quasars are even greater, and quasars are thus among the farthest known objects in the Metagalaxy.

Despite the extraordinary distance of the quasars from the Sun, some of them are identified with optical objects which are not exceedingly faint. Thus, 3C 273 is an object of stellar magnitude 12.6. The Sun would have the same magnitude from a distance of 1300 light years. Since the quasar is a factor of $1.3 \cdot 10^6$ more distant, the energy it radiates in the optical spectrum is a factor of $1.7 \cdot 10^{12}$ greater than the Sun's radiant energy, reaching about 10^{46} erg/sec. The amount of energy radiated by 3C 273 in the infrared region of the spectrum is one order of magnitude higher than the optical radiation. The total radiation power therefore may exceed $3 \cdot 10^{47}$ erg/sec. Of this, 10^{44} erg/sec are radiated in the radio spectrum. An estimate of the same order of magnitude is obtained for the quasar 3C 48, which is two and a half times as distant as 3C 273.

Comparing the optical radiation power of a quasar with the corresponding figure for the Galaxy (about 10^{44} erg/sec) we see that quasars emit two orders of magnitude more energy in the optical spectrum. Our star system is one of the giant galaxies. A normal average galaxy emits only 10^{43} erg/sec, i.e., one thousandth of the quasar power. Quasars are thus the most powerful sources of radiation in the Universe.

The dimensions of the quasars at first were expected to be proportionately large, but they are in fact much smaller than galaxies. This remarkable conclusion is based on direct measurements of the angular diameters from radio observations, which show that quasars are never larger than 0.1 of the diameter of an average galaxy.* Light variation of quasars suggests that their diameters are actually smaller, not more than a few light years. This light variation is a phenomenon of independent interest, and we will consider it in some detail.

Periodic photographic surveys of the sky have been now under way for more than half a century. They are used in solving various astronomical problems, and examination of a few thousand photographs of sky areas with the brightest quasars in them have shown that the brightness of quasars is markedly variable (approximately by a factor of 1.5 in 10—15 years). The light variation is periodic, and some data indicate the existence of periods as short as a few months and possibly weeks. However, even if

* Radio methods at this stage do not give a more exact determination.

we accept only the firmly established ten-year periodicity of the variations, the immediate conclusion is that the diameters of the quasars are surprisingly small, in relation to their tremendous energy output; they are tens of thousands of times less than the diameter of the Galaxy.

To understand how light variation leads to this conclusion, we have to recall one of the fundamental postulates of the theory of relativity, which states that no signal or interaction can propagate with a velocity exceeding the velocity of light. Since the brightness of the entire source varies in ten years, its dimensions cannot exceed 10 light years. If the source is larger, we have to assume that the different parts of the source change their brightness synchronously and in phase, without interacting with one another. This is a nonsensical assumption, since no part of the source "knows" what other parts distant more than 10 light years from it are doing.

We see that quasars are markedly different from the various types of star systems described so far, the main difference being their tremendous energy density. Quasars emit a hundred times more energy than the entire galaxy, and their volume is of the order of the combined volume of all the stars in the Galaxy.

The emission lines in the spectrum of quasars suggest that these objects are surrounded by extended regions of tenuous gas. The nebulae surrounding some of the quasars are actually visible on photographs. The emission lines appear superimposed on a very strong continuous spectrum. The continuous radiation originates in the quasar itself and the observed light variations are in fact associated with changes in the continuum intensity. The diameter of the nebulae is much greater than the diameter of the quasar, reaching a few hundred light years. Their volume is therefore also very large, and despite the low density of the gas in these nebulae the energy emitted in the spectral lines is only 10—100 times less than the radiation of all the stars in the Galaxy.

The mass of the gas in one such nebula is of the order of $10^6 M_\odot$. Judging from the great width of emission lines, the gas moves with tremendous velocities which reach 2000—3000 km/sec. However, even at these velocities, the gas does not disperse in space. It is apparently held together by the gravitation of the quasar, whose mass is therefore not less than $10^8 M_\odot$.

The nebulae around quasars are possibly formations of complex structure: they may be made up of individual dense jets or filaments, like the Crab Nebula. The radio emission of quasars is another factor stressing the possible likeness between these objects and supernova remnants (despite the obvious disparity in the scale of the phenomena). The spectrum of the radio waves and their polarization indicate the synchrotron origin of this radiation. The nebulae around quasars therefore contain a sufficient number of relativistic electrons which move in magnetic fields. The synchrotron mechanism is apparently also responsible for a certain fraction of the optical radiation of the quasar in the continuous spectrum, since the continuum radiation is partly polarized.

The radio source is much greater in diameter than the optical object. For example, the radio source of 3C 273 extends over more than a thousand light years. Estimates of the magnetic field H in the nebula from the radio spectrum of 3C 273 give $H \gtrsim 2 \cdot 10^{-3}$ oersted. Hence, the overall quantity of magnetic energy E_{magn} in the nebula (equal to $\frac{H^2}{8\pi}$ multiplied by the volume of the radio source) is of the order of 10^{57} erg and higher. The energy of relativistic electrons in the nebula reaches a comparable figure.

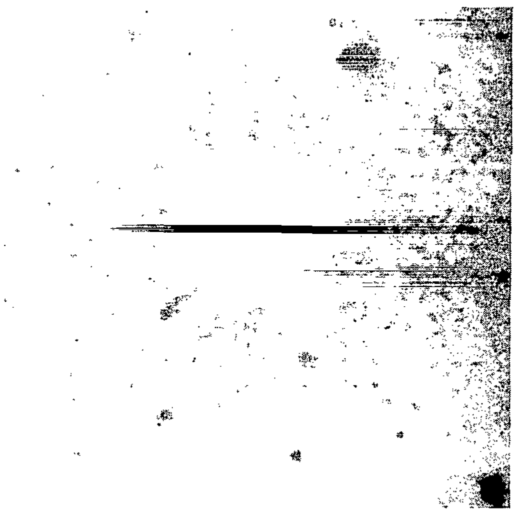

FIGURE 44. A photograph of the quasar 3C 273. Notice the ejection in bottom left corner. The large size of the "star" is associated with photographic effect and does not reflect its true size.

Judging from the structure of some quasars, their age is a few hundred thousand years. Thus, 3C 273 is a double radio source. The weaker radio component (B) coincides with the primary quasistellar source of optical radiation. The other radio source (A), which is identified with a weak optical object, is separated by 160,000 light years from B.* It is greatly elongated in the direction AB, and can be regarded as an ejection or a stream extending from B. If some time in the past A separated from B, the event occurred no less than $2 \cdot 10^5$ years ago, and possibly in a much more distant past, since A was hardly moving with a velocity close to the velocity of light at the time of the separation.

Photographs of the nebula around 3C 48 show streaks with a length of at least 200,000 light years. These streaks were apparently ejected from the quasar and it has thus existed in the active state at least for a few hundred thousand years.

It thus seems that the quasars have been radiating at the present-day level (and possibly even at a higher power) for at least 10^{13} sec. In this time, a quasar could radiate $10^{60}-10^{61}$ erg. The combined energy of the magnetic fields and the relativistic particles clearly cannot sustain this

* The true separation AB may actually be greater than this figure, since we are observing only the projection of AB onto the celestial sphere.

steadily persisting radiation, even if it were converted into radiation with hundred percent efficiency. We are therefore faced again, and this time with greater acuteness, with the problem of possible energy sources sustaining such powerful radiation.

The unusual properties of quasars gave rise to numerous hypotheses purporting to explain the nature of these curious objects. However, the quasars still remain an enigma. We will nevertheless review some of the theories, since their failure may help to concentrate our attention on new approaches to the whole problem of quasars.

According to one of the theories, the quasars are produced by a large number of supernova explosions in a peculiar object with a very high concentration of densely scattered stars. If the explosion of one star triggers explosions in its nearby neighbors, we have a "chain reaction," but whether or not such a process is physically permissible remains undecided. Moreover, the observed structure of quasars, and especially of 3C 273, is difficult to reconcile with this theory, which no longer enjoys any popularity. In general, thermonuclear reactions are not a highly efficient source of energy, as only 1% of the total rest mass is released in these reactions. The assumption of thermonuclear reactions taking place in an object with a mass of $10^8-10^9\,M_\odot$ does not save this theory, since the overall pattern of thermonuclear energy release in objects of such tremendous mass is completely different from what we observe in quasars.

A much more efficient method of releasing large quantities of energy is through gravitational collapse. If the mass of a body is M, its gravitational collapse may release a maximum energy of $0.8\,Mc^2$. Since the quasar masses are greater than $10^8\,M_\odot = 2 \cdot 10^{41}$ g, the gravitational energy corresponding to one tenth of the total mass, say, is 10^{62} erg, which is quite sufficient for sustaining a quasar. The gravitational collapse theory have thus become increasingly popular in the recent year: it visualizes the quasar as an object of gigantic mass $10^8-10^9\,M_\odot$ which collapses toward the center. Let us consider some typical features of the gravitational collapse.

An object of sufficiently large mass (greater than $10^6\,M_\odot$) cannot remain stable, since the interior pressure is unable to withstand the weight of the overlying layers. Therefore, if an object of such a tremendous mass has been created *in some way, it will inevitably collapse*. As the radius contracts, the gravitational energy is released. Suppose for simplicity that originally the radius was very large compared to its final value R. Then according to (12) the quantity of energy released in the contraction of a body of mass M is approximately $\frac{GM^2}{R}$. Using this expression, we can readily calculate to what radius an object of mass $10^8 M_\odot$ should contract in order to release 10^{62} erg of energy. This radius R is equal to $3 \cdot 10^{13}$ cm.

In these energy calculations, we have used the law of universal gravitation, according to which the gravitational attraction of an object is inversely proportional to the square of the distance. In accordance with the general theory of relativity, Newton's law does not apply at very small distances from the gravitating mass. Certain corrections should be introduced when the distance r is small compared to the so called gravitational radius $R_k = \frac{2GM}{c^2}$. For $M = 10^8 M_\odot$, we have $R_k \approx 1.5 \cdot 10^{13}$ cm, which is close to the above estimate of R. Therefore, when the final radius is estimated from the total energy released, we must introduce the general-relativity correction to Newton's law. This, however, does not markedly affect the result for R, which remains close to R_k. Thus, in order for an object of mass $10^8 M_\odot$ to release 10^{62} erg through gravitational collapse, its radius must shrink nearly to R_k.

As the radius of a collapsing object approaches R_k, the properties of the radiation escaping from the object markedly change. In particular, equation (30) expressing the gravitational red shift of the photons is modified. General relativity leads to the relation

$$\frac{\Delta\nu_0}{\nu_0} = -\frac{-2\frac{GM}{Rc^2}}{\sqrt{1-\frac{2GM}{Rc^2}}+1} = -\frac{\frac{R_k}{R}}{\sqrt{1-\frac{R_k}{R}}+1}, \tag{33}$$

which for R much greater than R_k clearly reduces to equation (30). For $R = R_k$ we have from (33) $\Delta\nu_0 = -\nu_0$, or the frequency shift is equal to the frequency itself. Therefore no radiation will reach the distant observer at any frequency ν_0, since the entire energy of the photons is expended in overcoming the gravitational forces. A collapsing object thus becomes invisible as soon as it has shrunk to the size of the gravitational radius. Perhaps the collapse continues beyond this point, but the object has vanished as far as the physical universe is concerned.

The surface of the collapsing object moves with a certain acceleration, as a body in free fall in the field of gravity. As the radius approaches the critical value R_k, the velocity of the surface approaches that of light. For an observer on the surface of the body, the time to reach R_k depends on the density of the body. If, say, the density of the object is comparable to that of the Sun, the collapse time is less than an hour. Conversely, for an observer situated far from the surface, the velocity of the surface appears to decrease as the object approaches the critical radius R_k, and from his point of view an infinte time is required to reach this radius.

The difference in the flow of time near the surface of a collapsing object and far from it is another curious effect in general relativity. It is essentially responsible for the gravitational red shift. Indeed, if the surface emits radiation of frequency ν_0, the period of one oscillation is $\Delta t_0 = \frac{1}{\nu_0}$. The same radiation for a distant observer has the frequency ν, and the corresponding period is $\Delta t = \frac{1}{\nu}$. Since $\nu = \nu_0 \sqrt{1-\frac{2GM}{Rc^2}}$, we have $\Delta t = \frac{\Delta t_0}{\sqrt{1-\frac{2GM}{Rc^2}}}$. The time between successive events on the surface of a collapsing object is thus less for an observer which moves with the surface than for a distant observer. As R approaches R_k, the difference in the flow of time becomes more pronounced.

Although the collapse hypothesis in principle explains the quasars both in terms of energy and the time scale, this theory meets with serious difficulties. First note that the formation of the excessively massive objects capable of gravitational collapse seems extremely unlikely. For instance, the contracting body should spin faster, and the increase in rotational velocity will enventually break the body into fragments. Moreover, it seems that the collapse of a spherical body cannot supply the required amount of energy. Thus, the collapse theory of quasars is still only half-baked.

In conclusion of this section note that a new class of objects — so-called quasistellar galaxies — were discovered in 1965. The optical radiation of these objects is similar to that of quasars, but they significantly emit no radio waves. Further study of quasistellar galaxies will probably bring us closer to solving the riddle of quasars.

§ 10. THE ROLE OF EXPLOSIONS IN THE EVOLUTION OF STARS AND GALAXIES

The rapid improvement of the astrophysical techniques led to a certain explosion in the influx of new information about celestial bodies. The information that we obtain from astronomical observations is hidden in "coded" form in the various characteristics of the radiation from celestial bodies. This information is "decoded" using various physical theories, and the processes taking place in stars and nebulae are reconstructed in this way. The development of a scientifically consistent theory of structure of celestial bodies a highly complex undertaking, since we are dealing with processes which are largely inaccessible to direct observation, e.g., processes taking place in the stellar interior.

The study of the evolution of star and star systems is even a more complex problem. These objects have existed and evolved continuously for billions of years, wheras the period of astronomical observations covers only a few decades. Therefore, as new observational data are accumulated, a stage is reached when the current theories of stellar origin and evolution invariably fail and a new and more comprehensive theory has to be developed after a detailed revision of the whole problem.

Only a few years ago the majority of astronomers were of the opinion that stars and nebulae formed from tenuous diffuse matter by gradual compression and condensation. At present, however, this hypothesis has proved inconsistent. Insurmountable difficulties arose especially after the discovery of explosions in galactic nuclei.

Explosions in galactic nuclei are only one link in the sequence of cosmic explosions that we have described in this book. Table 2 is a summary of relevant data on various explosions.

The time scale in Table 2 gives the order of magnitude of the duration in which the consequences of an explosion are clearly visible. The time of the explosion itself is apparently always small. Even the powerful explosions of novae take no more than 10^4 sec.

As we see from Table 2 the range of intensity of cosmic explosions is very wide. The energy of the weakest and the strongest known explosions is in a ratio of more than 10^{30}. It is remarkable that despite this tremendous difference in the scale of explosions, there is considerable likeness in the processes accompanying the explosion. Cosmic explosions always release high-energy particles, in particular relativistic electrons. Moreover, all explosions are associated with magnetic fields; no synchrotron radiation is possible without magnetic fields, and both weak and strong explosions show this radiation in the radio and the optical spectrum.

TABLE 2

	Sun	Flare stars	Nova-like stars	Recurrent novae
Explosion energy (erg)	$10^{30}-10^{34}$	$10^{32}-10^{35}$	$10^{39}-10^{41}$	$10^{42}-10^{43}$
Time scale (sec)	10^{3}	$10^{3}-3\cdot10^{3}$	$10^{4}-10^{5}$	10^{7} 10^{8}
	Novae	**Supernovae**	**Galactic nuclei**	**Quasars**
Explosion energy (erg)	$10^{44}-10^{45}$	$10^{49}-10^{52}$	$10^{55}-10^{59}$	$10^{59}-10^{61}$
Time scale (sec)	$10^{8}-10^{9}$	$10^{11}-10^{12}$	$10^{13}-10^{14}$	?

Each explosion also ejects gases with velocities ranging from a few hundred to a few thousand km/sec. The energy released by the explosion is generally highly concentrated, and the explosion therefore greatly disturbs the surrounding medium.

Cosmic explosions are a very common effect in the Universe, and they apparently constitute an extreme manifestation of the tendency of stars and star systems to reach a state of maximum dispersion. Both matter and energy are dissipated through space, the former by mechanical ejection and the latter through radiation. The inverse of dissipation is the condensation of matter and energy, when matter originally occupying a large volume concentrates in a much smaller region of space. So far, the process of condensation of diffuse matter has not been observed on the scale of stars and galaxies. This does not mean that condensation never occurs, but the fact is that we can observe only dissipation.

Condensation can hardly lead to a rapid and explosive release of energy, since the condensation process encompasses a tremendous volume: after all the massive objects form in this way from highly rarefied gas. On the other hand, remember that the velocity of propagation of any interaction never exceeds the velocity of light.

How do cosmic explosions fit into a theory which regards stars and galaxies as forming from diffuse matter? According to this theory, stars are condensations of interstellar medium in a certain volume, and originally they appear as gigantic spheres of gas. The exact process describing the formation of these spheres from interstellar gas clouds is highly complex and we will not discuss it here. The process is determined by the various environmental conditions, such as the motion of the gas, magnetic fields, etc.

The gas from which the star condenses is initially assumed to be cold, and therefore the original temperature of the newborn gas is low. The gas pressure is corresponding low. Since the gas cannot withstand the gravitational attraction, the star will continue contracting, and the temperature in the stellar interior will rise as the potential energy is converted into heat. However, as long as the temperature has not reached a certain critical value, no thermonuclear reactions are triggered and the light from the compressing gas sphere is determined entirely by the conversion of potential energy into heat and then radiation.

The transformation of a cold gas sphere into a luminous star has never been actually observed, and this stage of stellar evolution is purely hypothetical. Calculations carried out for spheres of various masses show, however, how the brightness of the sphere will change with the contraction of its radius. The time of contraction is highly sensitive to the mass of the sphere. For spheres with masses much greater than the solar mass, the compression stage may take a few hundred thousand years, spheres of one solar mass will continue contracting for a few million years, and gas spheres with the mass of a red dwarf will take several billion years to complete their contraction.

When the temperature in the stellar interior has reached the critical value, nuclear reactions are triggered and further contraction ceases. The next stage in the evolution of the star has begun, and its light is sustained by the release of nuclear energy. Hydrogen is gradually converted into helium by thermonuclear reactions, which by assumption are restricted to the central core of the star only. The process of hydrogen burning has been described in the preceding. When the hydrogen reserve of the core has been exhausted the star contracts further and the temperature in its interior again increases. The contraction stops when helium burning reactions are initiated, and possibly reactions between other heavy nuclei.

Calculations based on this model show that the hydrogen burning stage could have ended so far only in highly massive stars (with masses greater than $1.5-2\,M_{\odot}$). Dwarf stars, on the other hand, expend their hydrogen fuel very slowly and they can go on emitting light at the expense of hydrogen being converted into helium for a duration of time which is longer than the age of the Galaxy. These stars are therefore evolving without any marked changes in structure or luminosity. This is indeed the case of the Sun. Geological and paleontological findings seem to indicate that the solar radiation has not changed much during the last few billions of years.

The evolution of a star during the nuclear burning stage proceeds quietly, without any interruptions, and it is only at the very last stage of evolution of a giant star that a certain fraction of stellar matter may be detached and propelled into space. This is not necessarily an explosive process, however. The mass is probably lost by a sort of steady and continuous ejection from the parent star.

Note that this fragmentary evolutionary theory has been developed for solitary stars only. The interactions between the stars in binary systems can effectively modify the course of evolution by including strong motions in the stellar matter.

We have seen that stellar explosions occur in low-luminosity and low-mass stars (excepting type II supernovae), and the theory of stellar evolution does not predict an inevitable explosion for these stars. Explosions within the framework of the condensation theory are thus regarded as a secondary process, not related to specific evolutionary trends. The binary character of stars possibly plays a certain role in cosmic explosions, but we know it for a fact that solitary stars also explode.

Although explosive processes seem to be unrelated to the evolution of a star formed from a gaseous cloud, they may affect the evolution in certain cases. The exploding star loses some of its mass. Thus, if the mass of the envelope stripped off the star is greater than the solar mass (which is apparently in the case for type II supernova), the internal structure of the star completely changes.

Not only stars but also galaxies and even clusters of galaxies are assumed to have formed from diffuse matter. We have already noted that the intergalactic gas is exceedingly rarefied, and that the bulk of the matter in the Metagalaxy is apparently concentrated in stars. In the distant past, when the galaxies formed from diffuse matter, its density was apparently much higher.

There are two important factors corroborating the theory of condensation of stars and galaxies from diffuse matter. First, agglomerations of gas and dust occur in parts populated by what we call young stars. These young stars could not have drifted far from their place of origin. On the other hand, numerous solitary stars, binary systems, and multiple star systems, in particular galaxies, possess a high angular momentum.* We know from mechanics that the angular momentum of any isolated system cannot change without the action of external forces. This is one of the fundamental conservation principles, on par with the principle of energy conservation. In a large mass of randomly moving gas, there are always some clouds with a considerable total angular momentum. Therefore, if clouds and galaxies have formed by condensation of diffuse matter, we have an obvious explanation of the observed high angular momenta of stars and star systems.

The condensation hypothesis was first advanced as early as the 18th century by the German philosopher Kant and the French mathematician Laplace, who assumed that the solar system originated by condensation of a primordial nebula. This hypothesis remained popular for a fairly long time. At the beginning of the 20th century Jeans suggested that the stars also formed by condensation of interstellar matter in the spiral arms of the Galaxy. According to Jeans, the central parts of galaxies are made up of gas, where the spiral arms have differentiated into stars. The theory naturally had to be revised when the central parts of galaxies had been shown to consist of individual stars as well.

The current theory of galactic evolution essentially amounts to the following. Some 10—15 billion years ago, our Galaxy was a cloud of randomly moving gas. In terms of chemical composition, hydrogen was the prevailing element. Various assumptions are made which are intended to explain how the gas has lost its kinetic energy and how it collected into stars and star clusters in various parts of the Galaxy.

High-luminosity hot (young) stars are located near the galactic plane, in the spiral arms which are rich in gas and dust. The stars condense from the interstellar gas with a low content of heavy elements, and in the

* We recall that the angular momentum is defined as the product of the rotation velocity by the mass of the body and the distance from the rotation axis. The angular momentum of a system of many bodies is the sum of these products for each individual constituent.

course of their evolution they convert hydrogen into helium and other heavy elements. The stellar matter is thus gradually enriched with heavy elements. Gases are ejected from all stars with varying power. The ejected stellar gas therefore injects heavy elements into the interstellar medium, and stars forming at a later stage from the interstellar gas will have a higher content of heavy elements than the early stars.

This argument explains the peculiar features of stars of different ages. The age of a star can be fixed with fair certainty for members of star clusters. An open cluster is a fairly short-lived formation, which has existed for no more than 1—2 billion years. Open clusters gradually dissipate among the other stars, mainly due to the attraction by outside masses. The very existence of an open cluster is therefore an indication that its constituent stars are all relatively young, much younger than the Galaxy. All the stars of a cluster clearly formed at the same time: there is absolutely no other way to explain the clustering of hundreds of stars showing a common motion and other common features. Globular clusters, on the other hand, are very compact and they break up much more slowly than open clusters. They are thus probably much older objects than the open clusters. The fit between the observed features of globular and open clusters and the theoretical results is a tentative confirmation of the above theory of nuclear evolution of stars. This, however, has no bearing on the origin of the Galaxy itself.

The initial stage of formation of galaxies from the diffuse medium is even less accessible to theoretical treatment than the formation of stars. Here we will not discuss the various possible processes of this formation, and only note in passing that none of the current theories predicts the strong explosions observed in galactic nuclei.

We have already noted the present-day tendency of stars and star systems to disperse. Transition from condense forms of matter to states of higher dispersion is constantly observed in all parts of the Universe. Besides explosions, which constitute the extreme form of dissipative processes, mass and energy steadily escape from stars in a more peaceful manner. Analysis of various observed dissipation processes led to an entirely different hypothesis, which regards the formation of stars and galaxies as an outcome of the transition of matter from some superdense state into observable forms of matter and energy.

Evidence in favor of this approach was supplied by a study of star systems in the Galaxy and of clusters of galaxies in the Metagalaxy. Let us start with the galactic star clusters. We have already noted that clusters cannot form as a result of the agglomeration of stars of different ages. Stars in clusters and in double or multiple systems are objects of the same age and the same origin.

Some twenty years ago new star systems, known as stellar associations, were discovered. The age of these stellar associations is a few million years, which is negligibly small compared to the age of the Galaxy. It has thus been established that star formation processes continue in the galaxy to this day.

Another type of very young star systems are the so-called Trapezium-type multiple stars. In ordinary multiple systems, two out of every three stars are relatively close to each other, and the third star is tens of times more distant. In Trapezium systems, on the other hand, the distances from any star to all the rest are always comparable. There are fairly few Trapeziums in the sky, and they are always made up of high-luminosity hot stars, just like the associations.

FIGURE 45. The Trapezium system in the central part of the bright Orion nebula.

Theoretical study of motions in Trapezium systems shows that these formations will break up in 1—2 million years. Any star is retained in a multiple system only by the gravitational pull of the other members, and it will escape as soon as it acquires a sufficient kinetic energy to overcome these attractive forces. The stars in Trapezium systems are prone to very close encounters in the course of several circuits around the center of gravity of the system. In such a close encounter, one of the stars may acquire excessive velocity and its kinetic energy will reach the super-critical value needed to escape from the system.* Calculations based on the observed masses and motions of stars in Trapezium systems gave an estimate of about a million years for a critical encounter between stars. Hence the very existence of Trapezium systems is a proof that the constituent stars formed all together and fairly recently. These systems definitely cannot form as a result of random encounters between solitary stars. All the stars in each of these systems were apparently part of some massive dense body, which broke up into separate components.

Galaxies also occur in multiple systems. There are double galaxies, systems of several galaxies, and even clusters comprising thousands of galaxies. Multiple galaxies are fairly frequent, and they also occur in Trapezium-type systems, as well as normal groups. The same considerations as before show that the galaxies in Trapezium systems apparently originated from one massive object. The velocities of galaxies in a multiple system are generally such that a galaxy completes one revolution in about a billion years, which is not too small compared to the age of galaxies. Therefore, most of the galaxies simply have not had any chance

* In ordinary multiple systems, the motion of stars is governed by Kepler's laws, so that no close encounters are possible and the systems remain stable, at least over billions of years.

for close encounters so far, and this explains why numerous multiple systems are still observed intact.

In some cases we are actually witnessing the breakup of some Trapezium systems. An excellent example is the so-called Stephan's Quintet, a system of five galaxies. The velocity of one of the galaxies in this system is more than 1000 km/sec faster than the velocities of the other Quintet members, whose velocities all coincide within 100 km/sec. To retain a galaxy moving with a relative velocity of 1000 km/sec, the masses of the other galaxies in the system should be much greater than those observed in the Quintet. We are thus witnessing the actual escape of a galaxy from the Quintet. Other examples of dissipating systems of galaxies are also known.

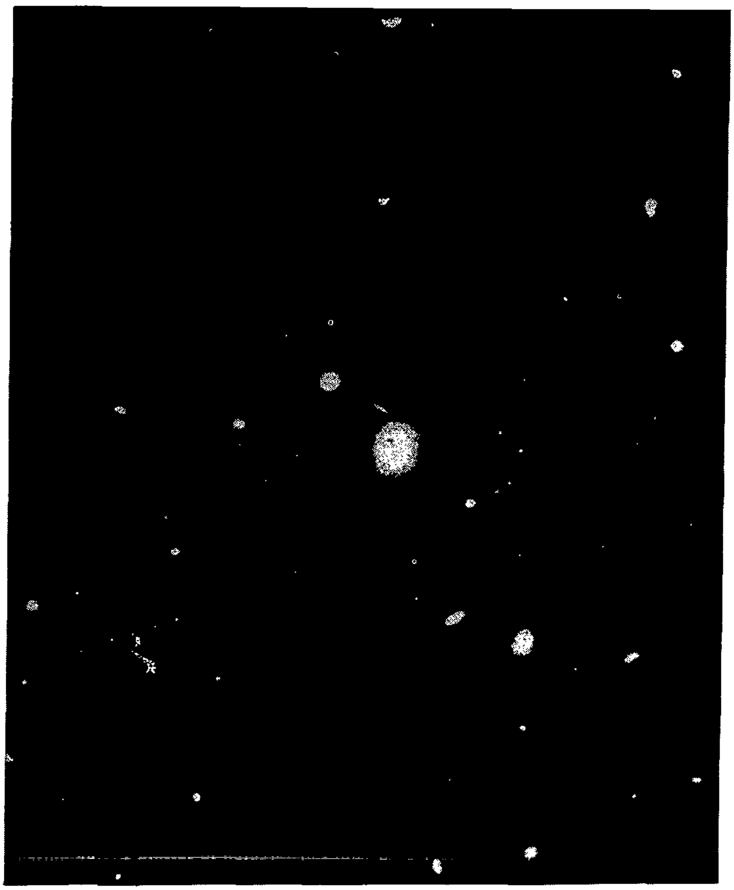

FIGURE 46. A cluster of galaxies in Coma Berenices.

In double galaxies, and in more complex multiple systems, luminous "bridges" are sometimes formed, linking two neighbors to one another.

This is still another proof of the close bond between galaxies in one system, which is probably associated with their common origin.

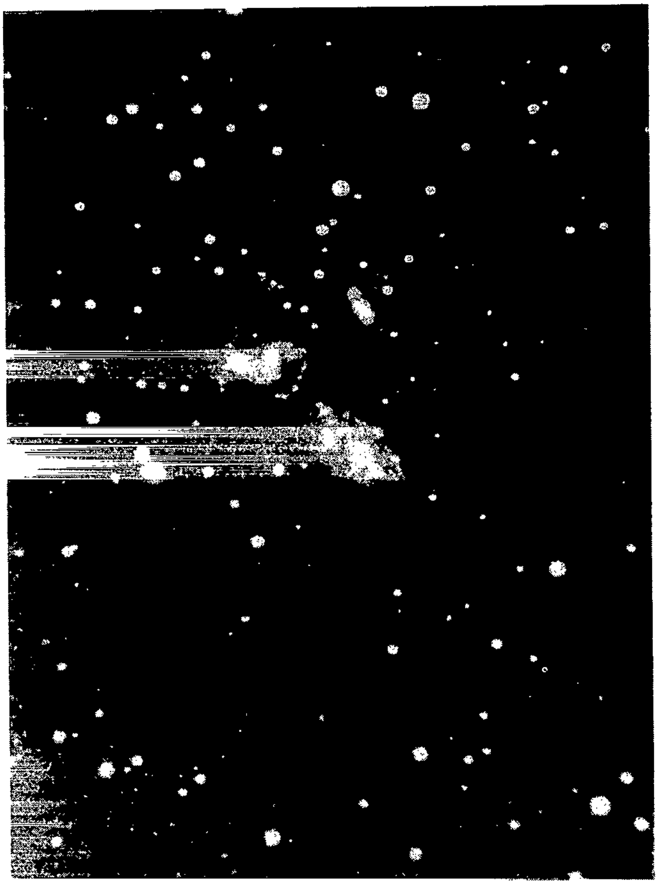

FIGURE 47. A Trapezium-type system of galaxies, Stephan's Quintet.

The data on hand thus indicate that giant star systems, as well as stars, apparently form when some dense massive bodies break up. This hypothesis also explains the nature of such "double galaxies" as Cygnus A, Centaurus A, and others. In these cases we are apparently dealing with two galaxies which formed very recently from a common source and they have not yet quite separated from each other. Since both galaxies move away from each other with tremendous velocities, and thus with extremely high kinetic energies, they were apparently formed in an explosive process. The powerful radio emission of the galaxies is a consequence of this explosion, which apparently generated relativistic particles in large numbers.

Observations of galaxies thus constantly provide new evidence in favor of the theories which associate the formation of galaxies with the breakup of other denser and more massive bodies. The evolution of a newborn galaxy leads to the formation of individual stars. As we already know, explosions in galactic nuclei in some cases lead to ejection of matter in a certain direction. It may be that the spiral arms of a galaxy are produced by such ejections from an exploding nucleus. A small "satellite" galaxy may also be ejected in this way from the nucleus.

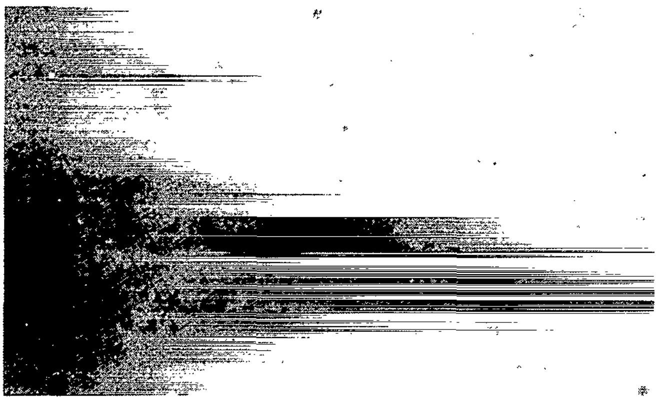

FIGURE 48. Two galaxies joined by a bridge (negative).

Our Galaxy also has two satellites: these are irregular galaxies known as the Large and the Small Magellanic Clouds. There are grounds to believe that the Magellanic Clouds are linked with the galaxy by extended bridges. The end of a spiral arm of the galaxy M 51, for instance, joins the main galaxy to a dwarf satellite galaxy. A similar phenomenon is observed in other systems, in particular in Andromeda Nebula. The bridges extending between double galaxies are also a modification of the same phenomenon. These bridges often appear to be extensions of the spiral arms. It therefore seems probable that the spiral arms of galaxies have evolved from such bridges between two systems which formed from a single body. If this is so, the formation of spiral arms can be connected with explosive processes in galactic nuclei.

This conclusion clearly is not final. As new observational data on galaxies become available, another more efficient mechanism of formation of spiral arms may emerge. However, the presence of jets and ejections reminiscent of spiral arms have been firmly established in some cases.

The breakup hypothesis at a first glance is inconsistent with the observed large angular momenta of galaxies: after all, a small-sized body cannot have a large angular momentum. The principle of angular momentum conservation can be satisfied if we assume that the two fragments forming from the original dense object spin in opposite directions. Their angular momenta are therefore of opposite sign, and although the angular momentum of each component is large, the sum of the angular momenta is adequately small. Observations of double galaxies, which apparently originated from the same source, indeed show that the spiral arms of the two components trail in opposite directions. This may indicate that the galaxies spin in opposite senses.

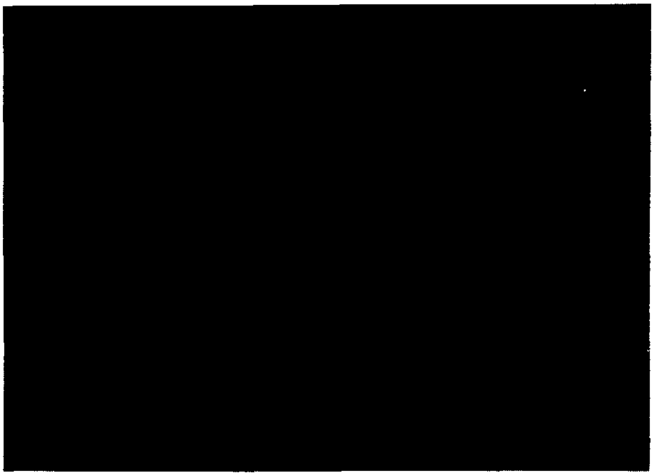

FIGURE 49. Magellanic Clouds, the satellites of our Galaxy: left, the Large Magellanic Cloud, right, the Small Magellanic Cloud.

According to this hypothesis, the stars form in the spiral arms which are ejected from the nucleus and thus contain exceedingly dense and massive objects. Fragmentation of these objects produces stars and also diffuse matter, gas and sut. The existence of hot (young) stars in the dusty regions of the Galaxy therefore cannot be regarded as a serious argument against the breakup hypothesis. We can equally well maintain that the existence of young stars and gas in the same region is indicatory of their common origin.

Neither the nature of the protostellar objects nor the formation of stars and diffuse matter from these objects have been clarified so far. However, the exsitence of stellar associations definitely shows that star-formation processes continue in the Galaxy to this day, although the spiral arms apparently formed in the distant past. It therefore seems that stars do not form right away from the "protostars" ejected from the nucleus: some time passes before the protostars break up into normal stars. At this stage we know absolutely nothing about the protostars. These are probably exceedingly dense and nonluminous objects.

The transition from protostars to another form of matter — stars — is probably an explosive process. This is evident in some cases from the high velocities of the stars in stellar associations, which cannot be accounted for by the collision mechanism. So far no such explosions have been observed, however, which is probably due to their extreme rarity.

Comparing the number of stellar associations in the Galaxy (with allowance for their lifetime from formation to dissipation) with the total number of stars in the Galaxy we conclude that an association forms once every 1000 years. This estimate assumes constant rate of star formation and regards all stars as forming in associations. More accurate calculations show that stellar associations in the neighborhood of the Sun (within a distance of 10,000 light years, say) form once every 100,000 years, while the observational period covers only a few decades.

Such cosmic explosions as the large flares of UV Ceti stars and T Tauri stars are apparently not related to star formation processes. Judging from the nature of these flares, however they can be regarded as a direct continuation of star-formation processes. Within the framework of these theories of star formation, flares are produced by blobs of protostellar matter rising from the interior to the surface.

The protomatter hypothesis of formation of stars and galaxies is still highly tentative, as we know nothing about the structure of protomatter. Some fundamental laws of physics will probably have to be modified before we can satisfactorily answer all the problems that arise in connection with star-formation processes. There is, of course, always a possibility that new observational findings will force us to abandon this approach altogether. At this stage, however, the protomatter hypothesis seems to be more promising than the hypothesis which attributes the formation of stars and galaxies to condensation of diffuse matter. If celestial bodies are assumed to originate from condensing diffuse matter, cosmic explosions of various scales cannot be linked up with the main line of evolution. If, on the other hand, matter is assumed to evolve from high-density to low-density forms, abrupt transition from one form to another is a basic stage of evolution.

CONCLUSION

We have covered a wide range of cosmic phenomena which take place far from the Earth. Some readers have probably wondered occasionally what was the use of this lengthy discussion of abstruse matters. The importance of solar flares and their effects on terrestrial processes have been discussed in some detail. In particular, we have emphasized that in future advance prediction of flares is highly important for preventing unnecessary exposure of astronauts to the hazards of space. Explosions in other stars, and particularly explosions in galactic nuclei clearly have no direct and immediate effect on the Earth. However, these explosions present excellent examples of high-energy processes which we can never hope to duplicate in our laboratories. Therefore we simply cannot afford to ignore these "experiments" that nature conducts against our eyes on a gigantic scale.

It is well known that a prerequisite of technical progress is thorough understanding of the physical laws involved. In studying new, and often quite unexpected types of cosmic effects, we advance toward further improvement of the laws of physics. Note, for instance, that the study of the thermonuclear reactions, which constitute the most powerful energy source known to man, began with observations of the Sun and the stars, and it is only much later that these reactions were realized in the laboratory.

At this stage we can hardly hope to harness and utilize any of the energy sources that we encounter in the study of cosmic explosions, although some of the effects of these explosions apparently can be simulated and put to work for us. Anyhow, additional information and deeper understanding of the world we live in eventually must have a certain effect on our daily life, although the future applications are not always clear to start with.

Another question that may naturally arise is who are the people working in this field of research? We have intentionally avoided giving specific literature references in the next: the range of effects is unusually wide and covers a lot of ground in astrophysics, so that numerous scientific organizations and probably hundreds, if not thousands, of individual scientists are active in this field. Within the framework of this book, we naturally cannot hope to properly assess the contributions of the different scientists, and this was not the aim. Our intention was to summarize the established facts which are essential for understanding the wide matrix of explosive processes. Nevertheless, we will now briefly survey the major institutions active in this field. The individual scientists mentioned are mostly heads of research groups.

We will start with chromospheric flares. In the USSR extensive research is being done at the Crimean Astrophysical Observatory under

A. B. Severnyi. Flaring stars of various types are also actively studied in the Crimean and at the Byurakan Astrophysical Observatory.

Novae have been a source of considerable scientific attraction for several decades now. Significant research is being done in USA at the Mount Wilson and Palomar Observatory, where the largest optical telescopes are installed (the 100-in. and the 200-in. reflectors). These instruments are also used to study supernovae, explosions in galactic nuclei, and quasars, which are accessible only to the largest telescopes. Large telescopes are now being installed in some other observatories as well, but this will hardly detract from the unique importance of the Mount Wilson and Palomar Observatory. Theoretical research into the problem of novae is being done in Leningrad under V. V. Sobolev, in Crimea and Moscow under E. R. Mustel' and others, in the Astrophysical Institute in France under E. Schatzman, in England, in the USA, and in other countries.

Nebulae produced by explosions of galactic supernovae were studied both optically and by radio methods in the USSR, Australia, and USA. The polarization of the radiation of the Crab Nebula was measured in detail at the Leningrad University and in Byurakan, and the assumption of the synchrotron origin of this radiation was confirmed. The problems of synchrotron radiation in application to supernovae had been originally studied at the Lebedev Physical Institute of the USSR Academy of Sciences and at the Shternberg Astronomical Institute by V. L. Ginzburg, I. S. Shklovskii, and others.

Explosions in galactic nuclei and supernovae are being intensively studied at the Byurakan Observatory under V. A. Ambartsumyan and in the largest observatories in the USA (J. Greenstein, A. Sandage, M. Schmidt). Radio observations of these objects are being carried out in England (the Manchester group), the USA, Australia, and other countries. Theoretical work on supernovae is being actively done in the USSR (under Ya. B. Zel'dovich at the USSR Academy of Sciences), in England (F. Hoyle with co-workers) and in the USA (Burbidge and Burbidge).

In the last section we have considered the two points of view on the formation of stars. The condensation hypothesis has been developed by various authors in various countries; considerable contributions were made by J. Oort (Netherlands) and L. Spitzer (USA). The diametrically opposite conception was advanced by V. A. Ambartsumyan. His hypothesis is based on a synthesis of results obtained from the research of nonstationary stars, on the one hand, and the study of stellar associations that he has discovered, on the other.

With these brief historical remarks we end our discussion of cosmic explosions. Following is a short list of literature for a closer study of the various topics considered in this book.

BIBLIOGRAPHY

Kompaneets, A. S. Udarnye volny (Shock Waves).— Fizmatgiz. 1963.

Frank-Kamenetskii, D. A. Plazma—chetvertoe sostoyanie veshchestva (Plasma—the Fourth State of Matter).— Atomizdat. 1963.

Goldberg, L. and L. Aller. Atoms, Stars and Galaxies. [Russian translation. 1948.] (The authors consider in detail, though in a popular style, a number of topics discussed in §2.)

Menzel, D. Our Sun. [Russian translation. 1963.]

Pikel'ner, S. B. Fizika mezhzvezdnoi sredy (Physics of the Interstellar Medium).— AN SSSR. 1959.

Kaplan, S. A. Fizika zvezd (Physics of Stars).— Fizmatgiz. 1961.

Kaplan, S. A. Elementarnaya radioastronomiya (Elementary Radio Astronomy).— Nauka. 1966.

Struve, O., B. Linds, and E. Pillans. Elementary Astronomy. [Russian translation. 1964.]

Greenstein, J., H. Chu, and G. Narlikar. Quasars. [Russian translation. 1965.]

Agekyan, T. A. Zvezdy, galaktiki, Metagalaktika (Stars, Galaxies, and the Metagalaxy).— Nauka. 1966.

Readers with a more advanced scientific background will find the following books of interest (these books also contain detailed bibliography on the corresponding subjects):

Schwartzschild, M. Stellar Structure and Evolution. [Russian translation. 1961.]

Gorbatskii, V. G. and I. N. Minin. Nestatsionarnye zvezdy (Nonstationary Stars). — Fizmatgiz. 1963.

Kaplan, S. A. and S. B. Pikel'ner. Mezhzvezdnaya sreda (Interstellar Medium).— Fizmatgiz. 1963.

Shklovskii, I. S. Sverkhnovye zvezdy (Supernovae).— Nauka. 1966.

National Aeronautics and Space Administration
WASHINGTON, D. C. 20546

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

NASA TT F-559
TT 69-55022

5440

Printed in Jerusalem, Israel